Onset-to-Onset Temporal Eye-Voice Span as an Indicator of Cognitive Investment in Reading Aloud and Sight Translation: An Empirical Study Drawing on Eye-tracking and Audio-recording Data

HAO ZHOU

How to cite:

ZHOU, HAO (2019) Onset-to-Onset Temporal Eye-Voice Span as an Indicator of Cognitive Investment in Reading Aloud and Sight Translation: An Empirical Study Drawing on Eye-tracking and Audio-recording Data. Doctoral thesis, Durham University.

Onset-to-Onset Temporal Eye-Voice Span as an Indicator of Cognitive Investment in Reading Aloud and Sight Translation:
An Empirical Study Drawing on Eye-tracking and Audio-recording Data

Hao Zhou

Submitted in Accordance with the Requirements for the
Degree of Doctor of Philosophy

School of Modern Languages and Cultures

Durham University, UK

2019

Abstract

Eye-Voice Span (EVS) is “a measure of the amount of material or time by which the voice lags behind the eyes in oral reading” (Morton 1964: 347). Both spatial EVS and temporal EVS happen naturally and inevitably during reading-speaking processes (Gibson and Levin 1975). A considerable body of psycholinguistic evidence has suggested that both spatial EVS and temporal EVS are very common during Reading Aloud (RA) (e.g., Buswell 1920； Laubrock and Kliegl 2015). Recently, researchers have also found that temporal EVS occurs ubiquitously during Sight Translation (STR) (e.g., Dragsted, Hansen and Sørensen 2009; Zheng and Zhou 2018).

The aim of this study is to measure the dynamic temporal distance between human subjects’ reading input and speaking output, to gain a better understanding of the nature of temporal EVS during STR and RA, and then to use it as an indicator of cognitive effort. Three groups of subjects from Durham University were recruited to perform RA and STR tasks in an eye-tracking laboratory. The RA and STR processes were recorded using a Tobii eye-tracker and an audio recorder.

The findings show the following results. Firstly, the subjects’ temporal EVS has a strong correlation coefficient with some major eye-tracking measurements, including total fixation duration, fixation count, and the sum of fixation and saccade durations. The present study, therefore, suggests that temporal EVS and these eye-tracking measurements are all representative of sub-categories of the cognitive effort devoted to completing a RA/STR task. Secondly, temporal EVS at sentence initials is found to be longer than temporal EVS at sentence terminals, as is spatial EVS. This finding indicates that the temporal EVS is not a random measurement, but one that varies predictably and potentially changes due to influences from cognitive processing. Thirdly, the duration of temporal EVS has a positive correlation with the cognitive effort devoted to the tasks, meaning that temporal EVS can serve as a dynamic indicator to measure the cost of cognitive effort in reading-speaking processes. Finally, temporal EVS is used to compare different STR processes, which provide potential applications for temporal EVS in future process-oriented translation studies.

Keywords: Eye-Voice Span, eye-tracking, sight translation, reading aloud.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AOI	Area of Interest
CAPS	Collaborative Activation-based Production System
cm	Centimetre
EVS	Eye-Voice Span
FC	Fixation Count
FFDur	First Fixation Duration
fMRI	Functional Magnetic Resonance Imaging
FPDurS	First Pass Gaze Duration on ST Words
FSD	The Sum of Fixation and Saccade Durations
FSI	Fixation Speech Interval
FTS	Fixation Time on Source Text as a Percentage of Total Task Time
GSF/GFP	Gaze Sample to Fixation Percentage
GTS	Gaze Time on Screen
H_0	Null Hypothesis
Hz	Hertz
KHz	Kilohertz
L1	First Language
L2	Second Language ms Millisecond
μ	Mean
ME	Metaphorical Expression
MFD	Mean Fixation Duration
mm	Millimetre
ms	Millisecond
n	Number
PCPD	The Percentage of Pupil Dilation
PGA	Pure Gaze Activities
r	Pearson Correlation Coefficient
RA	Reading Aloud
RAbM	Reading Ahead beyond MEs
RA-C	Reading Aloud in Chinese
RA-E	Reading Aloud in English
RAiM	Reading Ahead into MEs
s	Second
SD	Standard Deviation
S-I	Sentence Initial
SL	Source Language
SLA	Second Language Acquisition
SPG	Saccade Duration as a Percentage of Pure Gaze Activities
ST	Source Text
S-T	Sentence Terminal
STR	Sight Translation
STR C-E	Chinese-to-English Sight Translation
STR E-C	English-to-Chinese Sight Translation
STR E-G	English-to-German Sight Translation
TAPs	Think-Aloud Protocols
TFD	Total Fixation Duration
TL	Target Language
TPR	Translation Process Research
TT	Target Text
USP	Time of the Unclassified Sample as a Percentage of Total Task Time

Declaration

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Statement of Copyright

Acknowledgement

First of all, I would like to acknowledge the help given to me in the School of Modern Languages and Cultures by my supervisors Dr. Binghan Zheng and Dr. Sergey Tyulenev, from whom I have received helpful critiques and constant advice. Secondly, I would like to thank Dr. Susan Frenk and Mr. Stefan Klidzia for supporting my work at St. Aidan’s College SCR. Last but not least, I want to thank all the participants and my colleagues for their support and friendship.

It has been a great journey with laughter, excitements, and achievements, as well as tears, disappointments, and defeats. I am very grateful for my family’s unreserved support, love, and understanding, without which I wouldn’t be able to finish my study. Thank you, Durham University. Thank you, St. Aidan’s College.

I am shaped by my past. I am creating my future.

Chapter 1. Introduction

1.1 Research Background

Eye-Voice Span (EVS) is “a measure of the amount of material or time by which the voice lags behind the eyes in oral reading” (Morton 1964: 347). Both spatial EVS and temporal EVS happens naturally and inevitably during reading-speaking processes (Gibson and Levin 1975). In the last century, a considerable body of psycholinguistic evidence has suggested that both spatial EVS and temporal EVS are very common during Reading Aloud (RA) (e.g., Buswell 1920; Morton 1964; Levin and Addis 1979). Recently, researchers have also found that temporal EVS occurs ubiquitously during Sight Translation (STR) (e.g., Dragsted, Hansen and Sørensen 2009; Zheng and Zhou 2018). The aim of the present study is to measure the dynamic temporal distance between human subjects’ reading input and speaking output, and then draw some conclusions about temporal EVS during both STR and RA.

The motivation behind this empirical investigation originally came from an empirical study (Zheng and Zhou 2018), in which 24 subjects sight translated a passage containing Metaphorical Expressions (MEs) from English into Chinese. Their study showed that all the subjects have a time lag between the action of the eyes and the action of the voice when planning for sight translating MEs.

Also, Zheng and Zhou’s (2018) study showed that in most cases the time taken to read ahead into the MEs is longer than that taken to read ahead beyond the MEs. This finding indicates that the duration of temporal EVS is affected by the local processing difficulty caused by the source text. Hence, to identify the relationship between the duration of an individual’s temporal EVS and his or her cognitive effort became the main motivation for the present study. Therefore, I hoped to find out more about the potential of EVS to act as an indicator of cognitive effort, and also to apply it to analysing STR and RA processes.

Given that the cognitive demands of STR on interpreters are no less than those of simultaneous and consecutive translation (Agrifoglio 2004: 43), since the beginning of this century, the process of STR has attracted the attention of many researchers (e.g., Agrifoglio 2003; Agrifoglio 2004;

Lambert 2004; Sampaio 2007; Dragsted and Hansen 2009). Sampaio (2007: 65) emphasised the importance of mastering STR, as it is "an essential component of professional competence for interpreters, and a most desirable asset for translators". The critical viewpoint of Dragsted and Hansen (2009) is that performing STR and composing a draft for subsequent revision with the help of speech recognition software can reduce the time spent on revision and then gives the next generation of translators "a competitive edge in an ever more demanding market, where the players are not only other human translators, but also fully automated translation solutions" (Dragsted and Hansen 2009: 602). From this perspective, it is crucial that valid and reliable indicators to study the STR processes are established.

Although researchers have studied temporal EVS during RA in detail, the number of empirical studies on temporal EVS during STR is very small. Even less effort has been made to compare temporal EVS during STR and RA processes. The present study, therefore, aims to provide a detailed characterisation of temporal EVS, which represents the buffering system in these two reading-speaking modes. In addition, it is hoped that ultimately this approach of applying temporal EVS to study the STR and RA processes will also be of use to other researchers interested in this topic.

1.2 Research Aims, Hypotheses, and Questions

The primary aim of the present study is to gain a better understanding of the nature of temporal EVS during STR and RA and then to apply it as an indicator of cognitive effort. Before investigating the relationship between temporal EVS and the cognitive processing that takes place during RA and STR, the plan was to apply several major eye-tracking indicators to establish the hierarchies of cognitive investment during the process of RA and STR tasks: pupil diameter, total task time (TTT), the sum of Fixation and Saccade Durations (FSD), Total Fixation Duration (TFD), and Fixation Count (FC). The following two hypotheses regarding the hierarchies of cognitive investment during RA and STR were formed:

a. Subjects with different L1s invest different amounts of cognitive effort in processing the same reading material during RA.
b. Subjects invest a different amount of cognitive effort during STR than

during RA.

Once the hierarchies of the investment of cognitive effort were established, they were used as the foundation for a further analysis of several eye-movement indicators and temporal EVS. Three unconventional eye-tracking measurements, namely Mean Fixation Duration (MFD), Fixation Time on Source text as a percentage of total task time (FTS), and Saccade duration as a Percentage of the sum of fixation and saccade durations (SPG), were evaluated based on the hierarchies.

The hypothesis regarding temporal EVS was that there is a relationship between the duration of temporal EVS and the amount of cognitive effort devoted to the reading-speaking process. This correlation, whether positive or negative, would allow researchers to use temporal EVS as an indicator of cognitive processing in future studies. Based on this main hypothesis, the following research questions were set:

1. Does temporal EVS correlate with major eye-tracking measurements, such as total fixation duration, fixation count, and the sum of fixation and saccade durations?
2. Is temporal EVS at sentence initials longer than temporal EVS at sentence terminals, as spatial EVS is?
3. What is the relationship between an individual's temporal EVS and the cognitive effort invested in RA and STR tasks?
4. What can we learn about temporal EVS when applying it to analyse different STR processes?

1.3 Theoretical Basis

EVS during RA, both spatial and temporal, were defined by Buswell (1920) and Levin and Addis (1979) as the distance that the eyes are ahead of the voice. It is usually measured by "time, letters, letter spaces, ems (a printer's measure), syllables, or words, which is the most common index" (Levin and Addis 1979: 1). Since it can be measured by both spatial and temporal measures, for approximately five decades researchers have used the term EVS to refer to both temporal latency and spatial distance. However, spatial distance and temporal duration should be distinguished.

Almost a hundred years ago, Buswell (1921), who was one of the first researchers to apply spatial EVS in investigating subjects' reading competence, proposed that a person's spatial EVS expands as his or her reading competence develops. On the other hand, in this century, Inhoff et al. (2011) found that the duration of temporal EVS was routinely down-regulated to make the eyes closer to the spoken word through two complementary regulation strategies: increasing individual fixation duration and programming regressions. In this manner, a long temporal EVS is *corrected* (Inhoff et al. 2011: 554). Laubrock and Kliegl (2015) also discovered that the duration of temporal EVS at the beginning of a fixation correlates positively with the likelihood of having regressions and refixations. Hence, they argued that programming regressions and refixations are both ways of regulating the duration of temporal EVS if it becomes too long. Thus, whereas a substantially larger *spatial* EVS should accompany an individual's increasing reading competence, an overly long *temporal* EVS should be reduced; so, the two types of EVS should be dealt with separately.

If researchers separate temporal EVS from spatial EVS, they will be able to see that temporal EVS falls into the category of the *latency measurement* that is used in eye-tracking studies. A latency measurement is "a measure of time delay, i.e., the time between the on- or offset of one event to the on- or offset of another" (Holmqvist *et al*. 2011: 428). It is Holmqvist *et al*. (2011) who separated temporal EVS from spatial EVS, rebranded it as Eye-Voice Latency (EVL), and made it one of the latency measurements for eye-tracking research. To be clear, temporal EVS and EVL are the same thing, measured by a temporal unit (millisecond, to be specific). In the view of Inhoff *et al.* (2011: 543), temporal EVS embodies "a mechanism that maintains a close linkage between the identification and articulation of words through continuous oculomotor adjustments". This idea suggests that temporal EVS represents the cognitive binding of reading and speaking. In the present research, an attempt is made to distinguish temporal EVS from spatial EVS, with the focus being on the former in the eye-tracking study.

Nevertheless, spatial EVS was still very important for the research background. It has a long history in educational and psychological research (Levin and Addis 1979: 1), as well as in reading studies (Downing and Leong 1982: 145). In the early 20th century, spatial EVS during RA received a great deal of attention and substantial coverage as an intriguing topic in its own right. Later, researchers started to consider the relationship between spatial EVS and readers' reading skills, and the relationship between spatial EVS and grammatical characteristics of the reading materials. In the 1960s, psychologists' and educators' interest shifted to understanding the process of reading through studying the readers' spatial EVS (Levin and Addis 1979: 2), and spatial EVS has become an indicator of reading competence to some extent. Buswell (1920), who discovered a marked correlation between the length of the spatial EVS and the position in a sentence, stated that the subjects' average spatial EVS is the longest at the sentence initials and the shortest at the sentence terminals. Based on this finding, the present study hypothesises that although temporal EVS is different from spatial EVS, it might also have a similar pattern within sentences during RA and STR because both spatial and temporal EVS measurements are two manifestations of the same reading behaviour. Therefore, I consider whether the duration of temporal EVS also has a connection with the cognitive processing of sentences, and whether the duration of temporal EVS correlates positively or negatively with the amount of the cognitive effort invested in the STR and the RA tasks.

1.4 Methodology and Data Acquisition

To answer the research questions and to achieve the research aims, a series of experiments were designed and carried out. Three groups of voluntary and consenting subjects from Durham University were recruited. Three groups of subjects from three different countries were recruited to perform different RA and STR tasks, including Reading Aloud in Chinese (RA-C), Reading Aloud in English (RA-E), Chinese-to-English Sight Translation (STR C-E), English-to-Chinese Sight Translation (STR E-C), and English-to-German Sight Translation (STR E-G).

When Holmqvist *et al.* (2011) explained the use of EVL in a comprehensive guide book about eye-tracking methods and measures, they emphasised that EVL (temporal EVS) is the measure of the duration between the onset of a speech event and the onset of an eye-movement event. The present research, therefore, draws on experimental data combining eye-tracking and audio-recording. A Tobii eye-tracker and an audio recorder were used to record all the reading and speaking processes. The Tobii Studio and the Audacity software analyse the eye-movement data and the audio data. As an influential quantitative and qualitative research method, the eye-tracking method has provided extremely valuable information in reading studies (e.g., Rayner 1998; Jakobsen and Jensen 2008; Pellicer-Sánchez 2016). Meanwhile, eye-tracking has also been considered a well-established approach to investigate STR process (Dragsted and Hansen 2009; Shreve, Lacruz, and Angelone 2010; Korpal 2012; Chmiel and Mazur 2013; Zheng and Zhou 2018).

Previous studies on the spatial EVS and temporal EVS during reading aloud have used different units for calculation, including sentences, specific grammatical structures within a sentence, phrases, and words. The present study uses reading/translation units at sentence initials and sentence terminals for measuring temporal EVS at the sentence boundaries in both the RA and the STR tasks. Since Holmqvist *et al*. (2011), who branded temporal EVS as a latency measure for eye-tracking studies, stated clearly that "in the majority of studies, latency is counted from onset to onset" (Holmqvist *et al*. 2011: 443), the present study will follow their guidance and adopt this onset-to-onset algorithm. Temporal EVS will be measured by the duration between the onset of sound and the onset of eye movements. Figure 1.1 is an illustration of how onset-to-onset calculation is carried out.

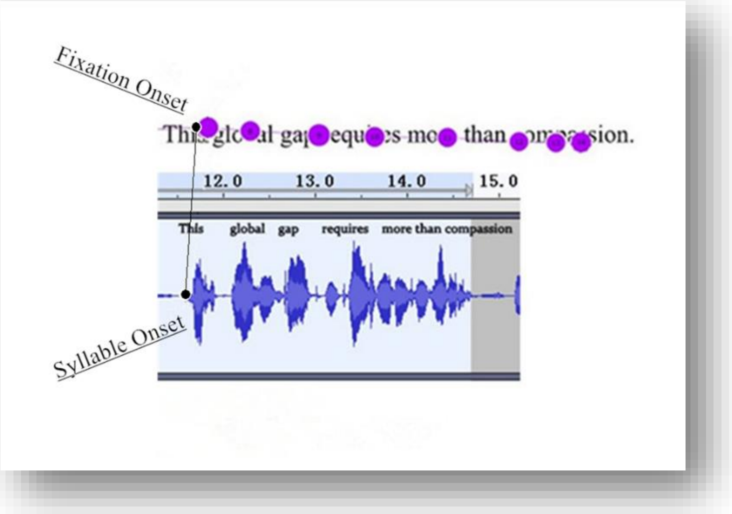


Figure 1.4a: Measuring Onset-to-onset Temporal EVS

In summary, onset-to-onset temporal EVS is the time span from the beginning of a fixation on a reading/translation unit to the onset of the pronunciation of the corresponding unit (time of the onset of articulation - time of the onset of fixation). The present research, in addition, proposes using one algorithm throughout the research and in any follow-up studies so that the results will be comparable.

1.5 Delimitation

The present research, drawing on eye-tracking and audio-recording, does not combine quantitative data with qualitative data. When studying temporal EVS during STR and RA, neither concurrent nor retrospective verbalisation can be compared to the two main data elicitation methods as carrying out concurrent verbalisation is not realistic during RA or STR. In addition, the pilot study revealed that retrospective verbal data did not yield enough findings regarding temporal EVS. Therefore, when studying temporal EVS, it might be of greater value to triangulate the eye-tracking and audio-recording methods with other online methods, such as a future special speech recognition programme.

Moreover, the present study is process-oriented, focusing primarily on the temporal EVS and cognitive processing during RA and STR. Analysing the relationship between temporal EVS and the products of the RA/STR tasks, however, is not within the scope of this thesis because it cannot be assumed that there is an absolute positive correlation between the quality of the product and the amount of cognitive effort invested in the process.

Finally, like the previous studies on spatial EVS and temporal EVS, this study is not based on a longitudinal study. The present study hopes to lead to multiple hypotheses regarding the development of temporal EVS by comparing different test groups. The findings and conclusions, hopefully, will be evaluated in longitudinal studies in the future.

1.6 Structure of the Thesis

Chapter 1 is an introduction to this empirical study. It elaborates on the research background, motivation, aims, hypotheses, and research questions. Moreover, it gives a brief account of the theoretical basis, methodology, data acquisition, and delimitation. In the last section, the structure of the thesis is presented.

Chapter 2 and Chapter 3 establish the theoretical background of the thesis by reviewing previous research undertaken that relates to the STR process and temporal EVS in reading-speaking modalities. Chapter 2 reviews the empirical studies that have investigated the STR process and gives an overview of the related topics, including the definition of STR, STR varieties, STR problems and difficulties, STR competence, its similarities and differences with RA. Chapter 3 sets up the theoretical framework. It elaborates on the important notion of the both spatial and temporal EVS during RA and STR. Then, it revisits this topic from cognitive perspectives, proposing that temporal EVS has the potential to become an eye-tracking indicator of cognitive effort.

Chapter 4 and Chapter 5 outline the foundation of the research approach by discussing the research methodology first, and then presenting the way in which the data are collected. Chapter 4 begins by introducing the research methodology that consists of eye tracking and audio-recording. This chapter also introduces and discusses a series of eye-movements and corresponding eye-tracking indicators, including pupil diameter, fixation-based eye-tracking indicators, saccade-based eye-tracking indicators, and combined indicators. Chapter 5 sets out the research design by introducing the recruited subjects, stimulus, and experimental tasks. It also elaborates on the findings of the pilot study and the adjustments made based on it, as well as the final experimental procedures. This chapter also explains its experiment settings including the apparatus, fixation filter, segments, scenes, and Area of Interest (AOI). Furthermore, it gives a detailed account of EVS coding and calculation. Then, reviews the status of eye-tracking data quality assessment. The four data quality assessment criteria used by the present study are established and explained. They are: Fixation Time on Source text as a percentage of total task time (FTS), the time of the Unclassified Sample as a Percentage of total task time (USP), Mean Fixation Duration (MFD), and the standard range of the Saccade duration as a Percentage of pure Gaze activities (SPG). The chapter will conclude with a presentation of the results from the eye-tracking data quality assessment.

Chapter 6 serves to accomplish a preliminary research aim. First, it investigates the cognitive investment in each reading-speaking task through both inter- and intra-group comparisons, based on an analysis of a series of eye-tracking measurements. The indicators used in this chapter are pupil diameter, total task time (TTT), pure gaze activities (PGA) (also known as the sum of fixation and saccade durations, FSD), total fixation duration (TFD), and fixation count (FC). This chapter then investigates the reliability and validity of a measurement called Mean Fixation Duration (MFD). Moreover, two other potential indicators used as measurements of the eye-tracking data quality, namely the Fixation Time on Source text as a percentage of total task time (FTS) and the Saccade duration as a Percentage of pure Gaze activities (SPG), are explored and discussed.

Chapter 7 is dedicated to reaching the primary research aim. Temporal EVS during RA and STR is analysed in this chapter. The first section presents the results coming from the three groups. Then, this chapter answers the research questions through the following steps. Firstly, this chapter touches upon the relationship between temporal EVS and three major fixation-based indicators including total fixation duration, fixation count, and the sum of fixation and saccade duration. The aim is to see if temporal EVS correlates with these eye-tracking measurements. Secondly, it will closely examine temporal EVS in sentences. If there is a specific pattern within and across unit boundaries, it means temporal EVS is similar to spatial EVS in this aspect. Thirdly, the next step is to establish the relationship between the duration of temporal EVS and the cognitive events of the mind. In other words, the present study aims to understand whether the duration of one's temporal EVS correlates positively or negatively with the amount of the cognitive effort he/she invested in the reading-speaking process. Finally, this study puts temporal EVS into application, in the hope of improving our understanding of different STR processes: forward and backward STR (STR E-C vs. STR C-E), and STR from the same Source Language (SL) to different Target Languages (TL) (STR E-C vs. STR E-G).

Chapter 8 summarises the main findings of this study, revisits the topic of temporal EVS, discusses the limitations of the present research, and considers future avenues of research.

Chapter 2. Sight Translation (STR)

2.1 Sight Translation (STR): An Overview

Sight translation (STR), also known as “prima vista” translation (Jakobsen and Jensen 2008: 116), presents translators with a written Source Text (ST) and requires an immediate oral reformulation of the ST in the target language. It is “a specific type of written translation as well as a variant of oral interpretation” (Lambert 2004: 298). Nowadays, STR is not only exercised frequently in public service interpreting, but also included in handbooks and assessment of public service interpreting (Vargas-Urpi 2018). Besides that, this type of translation can take place in a variety of circumstances. For example, the interpreters might receive a script of the speaker’s speech during a conference to facilitate simultaneous interpreting with STR; they might be required to concurrently sight translate the content shown on the screen during a presentation in real-time; documents and presented exhibits might need to be sight translated immediately in court; doctors may need an accurate sight translation of foreign medical prescriptions; and tourist guides are always sight translating signs and introductions to tourist attractions. In addition, Nilsen and Monsrud (2015: 10) reported that interpreters in the public-sector services in Norway perform STR almost every day. Given the current use of STR in practice and teaching, I believe that STR will be an even stronger trend in the translation market and translator training in the future. Therefore, STR is worth investigating and hopefully this study will result in some progress towards understanding cognitive processing during this reading-speaking modality through the utilisation of an up-to-date research methodology.

As the hybrid modality between written translation and oral interpretation (Dragsted and Hansen 2009: 601), STR’s “cognitive demands on the interpreters are by no means less than those of simultaneous and consecutive translation” (Agrifoglio 2004: 43). Gile (1997), Dragsted and Hansen (2009) described the characteristics of the most common type of STR. Firstly, the interpreter has the ST that continues to be visually accessible, hence, no particular memory effort is involved. Also, the interpreters have more control over their speed of delivery. Moreover, since the words of the source language are accessible throughout the process, the risk of linguistic interference may be higher in STR than in other kinds of interpreting (Gile 1997: 203-4).

There are different varieties of STR. The most common are *STR Proper*, *Prepared STR*, *Consecutive STR*, *STR in Simultaneous Interpretation with Text*, and *STR in Consecutive Interpretation* (Sampaio 2007: 65). In general, *STR Proper* refers to an oral reformulation of a written text immediately after receiving it. *Prepared STR* differs from *STR Proper* in the sense that it gives the translators a certain amount of time to read the written text, consult the glossary, and search for background information in advance. During *Consecutive STR*, the translators are required to paraphrase and summarise the ST, instead of delivering every sentence in the ST as in *STR Proper*. *STR in Consecutive Interpretation*, on the other hand, is the mainstream consecutive interpreting task with text. *STR in Simultaneous Interpretation with Text*, also known as *STR Interpreting* or *Simultaneous Interpreting with text* (Gile 2009: 181), provides the interpreter with a written copy of the source content, either on screen or paper, during the performance of a simultaneous interpreting task (Sampaio 2007: 65). All these different types of STR belong to "a hybrid genre in that written text is read and transformed by the translator/interpreter into the spoken modality" (Jakobsen and Jensen 2008: 106).

The present study uses different levels of STR, which could happen based on different task complexity, requirements, and purposes. Level 1 STR is a relatively linear process as the translator verbalises the text one word/expression after another into the target language. Level 2 STR consists of larger translation units, within which the order of words might be different from that in the ST. The order of phrases and clauses, on the other hand, mostly abide by the sentence structure of Target Text (TT). Level 3 STR is more flexible than the previous levels, but is still a faithful transformation of the thoughts and ideas from the ST to the TT, meaning that all the written information is successfully transformed using correct grammar and lexis. Level 4 STR seeks a functional equivalence between the ST and the TT. The style and register of the original text should be maintained. At this level, translators may adopt a variety of translation techniques, such as omitting, elaborating, and summarising, to achieve the functional equivalence.

Progression from level 1 to a higher level or high levels could happen when the translator becomes more experienced and skilled. It could also take place one after another when a translator approaches the same task. In summary, moving from level 1 to level 4 of STR is a process starting with linguistic code-switching and changing to constructing an functionally equivalent TT.

Researchers hold different views regarding the nature of STR. For example, Moser-Mercer (1995) argues that since both visual and oral information processing is involved, sight translation could be viewed as either a special type of written translation or a variant of oral interpretation (Moser-Mercer 1995: 159). Many others have considered STR to be closer to interpreting than to written translation because of its time restriction and the oral nature of the task (Viezzi 1990; Martin 1993; Agrifoglio 2003; Agrifoglio 2004; Dragsted and Hansen 2009). Stansfield (2008) suggested that STR is a task that has more in common with oral interpreting than with written translation, and therefore is best performed by an interpreter. In many cases, researchers, translators, and translation teachers view STR as preparation or a training method for simultaneous interpreting lessons (Song 2010). In Kim’s (2001) research, when surveyed, translation students argued that STR is closer to oral interpreting, rather than to written translation. Its cognitive demands, however, are by no means less than those of the interpreting task (Agrifoglio 2004). Hence, some others see the potential for STR to become an independent subject and to be taught as an important skill to trainees, as it is valued in the translation workplace, in translator and interpreter training, and in language classes. For instance, in an attempt to explore the notion of “speaking your translation” (Dragsted and Hansen 2009: 601), Dragsted and Hansen (2009: 595) found that “the substantially lower output rate in the translators’ written translation compared with their STR does not seem to be justified by higher quality output”. Their finding supports Ericsson’s (2010) idea of incorporating STR as a deliberate practice activity in translation training curricula.

The similarities and differences between STR and other translation modalities have been discussed and analysed by many researchers. For instance, Gile (1984) built the *Effort Model* for STR, along with models for other interpreting modalities. After establishing the *Effort Model* for simultaneous interpreting (Simultaneous Interpreting = Listening and Analysis + Production + Memory + Coordination), Gile (2009) revised the *Effort Model* for STR as: STR = Reading + Short-term Memory + Production + Coordination. It is relatively obvious that the STR process starts with reading input and ends with a completion of oral output. It is not STR without either of these two efforts. However, below the surface, cognitive effort consumed by short-term memory and coordination is considerable due to the time limit and the natural flow of the STR process. In STR, the *Listening and Analysis Effort* is replaced by a *Reading Effort* as the task starts with reading the written ST and requires a spoken TT. The fundamental hypothesis of the *Effort Model* for STR, as well as for other oral interpreting modalities, is that an interpreter's performance would deteriorate when the cognitive demand is larger than the available cognitive effort (Gile 2009).

STR has been studied under a variety of conditions (Agrifoglio 2004; Lambert 2004; Setton and Motta 2007; Dragsted and Hansen 2009). For example, Dragsted and Hansen (2009) made a comparison between translators' behaviour during STR and written translation in a comparative study. The evaluation of the translation produced by professional translators and interpreters showed that there are fundamental behavioural differences between the interpreters and the translators. The finding that to some extent the forms of translation input and output influence translators' choices and strategies is another reason to distinguish STR from written translation and oral interpreting. There are also studies designed to investigate the differences in translators' behaviours when undertaking different types of STR processes. For example, Lambert (2004) compared subjects' performance in sight translation, sight interpretation (simultaneous interpretation *with* text), and simultaneous interpretation (without text). The experiment aimed at obtaining empirical evidence to examine whether the participants' performance is enhanced or hindered by the visual presence of the material. The result was that "visual exposure to the message to be interpreted does not necessarily *interfere* with a subject's already overloaded capacity to listen and speak simultaneously, but that in fact, it may even *help* the student's performance" (ibid.: 294).

Considering the role of STR in professional practice, it is essential to incorporate it as a component in any T&I curriculum (Sampaio 2007). In fact, according to Sampaio (2007: 64), some “well established and highly reputed interpretation programmes in the USA and Europe often have STR as a component of their entrance examinations in order to assess the linguistic and communicative potential of prospective students”. Besides examination, STR is also usually included in the curriculum of training programmes as sight translation skills help interpreting trainees react quickly during the translation process (Weber 1990). Many teachers of interpreting modules recommend regular STR exercises to their students because they consider oral translation skill an important component for the other interpretation modes. They believe that “the rapid and efficient visual-brain-vocal coordination required by STR constitutes the stepping stones towards consecutive and simultaneous interpretation” (Sampaio 2007: 64). Through practising the application of strategies, such as syntactically restructuring and paraphrasing, interpreting trainees can develop language transfer and speech delivery skills during STR (Ilg and Lambert 1996: 73). Therefore, a well-conducted STR course that “equates with an opportunity for the development and refinement of such oral translation skills” (Sampaio 2007: 65) should be included in translator/interpreter training programmes, “particularly for those who wish to maintain their professional standing and broaden their professional horizons” (ibid.). Furthermore, mastering sight translation skills is not only essential for translators and interpreters but is also helpful for language learners. Students’ bilingual competence, according to Sampaio (2007), is an asset in foreign language education, and STR exercises can be used to enhance both the learners’ First Language (L1) and foreign language proficiency.

Scholars such as Weber (1990), Moser-Mercer (1995), Lambert (2004), and Sampaio (2007), therefore, all highlighted the benefits of STR, and argued for it to be considered an ideal pedagogy for training translators, interpreters, and language learners. Others even suggest considering STR as an independent discipline due to its significant role in the translation workplace and its pedagogical implications (Sampaio 2007). For example, Al-Qurashi (2004) strove to investigate the problems that translators face in

English-to-Arabic STR. The study found that the subjects made different types of mistakes including lexical, syntactic, and extra-linguistic mistakes and, consequently, Al-Qurashi suggested that STR should be included as a separate course in the translation training curriculum. Stansfield (2008) discussed issues associated with sight translating the assessment content for English language learners, and pointed out that one must remember that not just any native speaker of a foreign language can perform sight translations. The ideal sight translator should be highly literate, be highly proficient in both languages, possesses rapid interpretation skills, and have a solid knowledge of both the source culture and the target culture (Stansfield 2008: 10).

As a result of the rapid development of technology and the increasingly demanding requirements from the clients, STR has become a new trend in the translation industry and has proved to be an efficient mode of translating used widely in conference interpreting and community interpreting (Li 2014: 68). At scientific conferences, STR tends to be applied more frequently because of the increasing number of non-native speakers, the downgrading of impromptu speaking skills, and the possibility of using visual aids in presentations (Weber 1990; Setton and Motta 2007; Gile 2009). According to Li (2014: 67), it was “agreed among the interpreting community that STR is an increasingly important practice, a valuable pedagogical tool, and a necessary component of learner needs”. Compared to written translation and oral interpreting, however, STR remains a relatively unexplored translation modality as far as academic research is concerned, and insufficient attention has been given to exploring its teaching methodology (Sampaio 2007: 63).

2.2 STR Problems and Difficulties

Translators must be aware of the potential problems and difficulties they will encounter and prepare themselves with potential solutions to perform a successful STR. STR, which requires reading and speaking in two different languages at the same time, is not an easy task, otherwise it would not have attracted interest from so many researchers and continue to be the research topic of present and future studies. Researchers, furthermore, have to know the potential problems with STR and difficulties when they design STR experiments to suit different research purposes.

For sight translators/interpreters, STR tasks involve a shift of their attention between a written input and an oral output. Compared with traditional written translation and oral interpreting, the shift itself may demand extra cognitive effort. Stansfield (2008) considered STR closer to oral interpreting and suggested that interpreters would have an advantage in performing STR over translators. This researcher argued that out of the three most common variations of interpreting, STR is the most difficult for the following reasons: Firstly, STR requires a shift from the written ST to the verbal target text (TT). This shift is unusual for interpreters who are used to listening to the ST and translating it into another language orally. Secondly, the written STs usually contain relatively complex syntactic structures, grammatical coordination, and vocabularies. Agrifoglio (2004), Shreve, Lacruz and Angelone (2010) also emphasised this reason, arguing that the information that needs to be extracted from all linguistic levels of a written ST during an STR task is usually much denser and more complex than that of the oral STs. Thirdly, Stansfield (2008) pointed out that having to use a different system of organisational cues poses problems for interpreters. Unlike translators who can use punctuation to indicate whether the present sentence ends and whether the present sentence is exclamatory, interpreters have to use pauses and intonation in their speech to deliver the same information. Finally, STR is particularly difficult because professional interpreters might perform it less frequently than other modes of interpreting.

Agrifoglio (2004) and Gile (1997) both discussed another STR problem: the risk of visual interference. Shreve, Lacruz, and Angelone (2010) suggested that due to the continued presence of the ST, sight translators might not perform deeper processing but repeat shallow scans of the linguistic input. As a result, written words become “a source of information noise obstructing the ability to easily extract meaning from the text” (Shreve, Lacruz, and Angelone 2010: 65). Their argument indicated that STR is extraordinarily sensitive to visual interference (Shreve, Lacruz, and Angelone 2010), meaning that the presence of the ST might actually hinder language output, rather than assisting it. Nonetheless, some interpreters might still find it helpful to check the previous content quickly during STR, especially when they forget what they have read.

The fact that the ST continues to be visually accessible to the translator (Agrifoglio 2004: 44; Gile 1997: 204) does not necessarily imply that “there is no memory effort” (Dragsted, Hansen, and Sørensen 2009: 293). During STR tasks with a time constraint, translators must hold some important information in their memory to avoid having to look back. To investigate the role of short-term memory during the production phase of STR, Pedersen and Dam (2014) conducted a quantitative analysis of the STR products. They reported that to produce and monitor their TT simultaneously, sight translators consulted their short-term memory “at least once to produce 71% of the target-text sentences selected for analysis” (Pedersen and Dam 2014: 103). This study also reveals a discrepancy between the subjects’ dependence on short-term memory during STR and their awareness of this matter. Pedersen and Dam (2014) thus suggested that Daniel Gile’s (2009) *Effort Model* for STR should be revised by either including the aspect of short-term memory during production in the existing *Memory* component that covers only the reading part of STR or adding a separate component that caters for the self-monitoring effort involved in STR processing (Pedersen and Dam 2014: 104). Since STR does not proceed in a strictly linear fashion, it requires a high level of self-monitoring during the process. Furthermore, an STR task requires effortful processing within a limited time. In fact, the time restriction is of great significance in a STR task. Pöchhacker (2004: 19) argued that if the translator/interpreter works without the constraints of real-time performance, STR will “shade into the consecutive mode or even come to resemble ‘oral translation’, with considerable opportunity for ‘reviewing’ and correction”. In other words, without time restrictions, STR is no longer STR. All the different types of STR, therefore, are incorporated with a “built-in speed norm” that automatically includes an element of time pressure (Jakobsen and Jensen 2008: 107). On the one hand, it is a distinctive feature of STR. On the other hand, it contributes to the difficulty of this translation modality to a large extent. Due to the time constraints of STR, the translators/interpreters have to comprehend ST input and produce TT output simultaneously (Gile 2009: 169-170). For the same reason, in the interpreting/STR models, translators tend to “typically settle on the first word that occurs to them” (Chafe and Danielewicz 1987: 88). Similar to performing simultaneous interpreting tasks, interpreters are not privileged with the time to revise their translation product during STR. STR interpreters, therefore, face a lot of difficulties and pressure although they might have more control over their speed of delivery during STR than during other interpreting tasks. Having only limited time to deliberate the ST and no opportunity to reconsider their choices even if they were not satisfied, sight translators/interpreters need to devote a significant amount of attention and cognitive resources to monitor the production process.

2.3 STR Competence

Since STR is viewed as a translation modality and research topic, many researchers have tried to investigate STR skills and competence, in the hope of helping translators improve their STR performance. Like STR problems and difficulties, STR competence also has the potential to influence one's cognitive processing during STR. In light of this potential influence, this section provides a review of STR competence. Here, we shall start with addressing translation competence, to which STR competence is closely related.

Research on translation competence started when developments in written translation studies in the mid-20th century led to a series of attempts to develop theories of translation and explore its pedagogy (Schäffner and Adab 2000). Translation competence and its acquisition form a significant link between translation research, practice, and training. For decades, researchers have been hoping to gauge what constitutes translation competence and how it can be developed through carrying out empirical translation studies. Because of its close relation with foreign language acquisition, translation competence often "appears as the automatic by-product of second-language competence" (Göpferich and Jääskeläinen 2009: 174). However, translation competence is not merely a bilingual competence. Mastering two or more languages is not the sole prerequisite for conducting translation or interpreting, nor is it a guarantee of a good translation/interpreting product.

What constitutes translation competence is a long-standing question that translation teachers have been trying to understand for decades (Göpferich and Jääskeläinen 2009: 174). Vienne (2000: 91) viewed translation competence as a straightforward three-level skill: de-verbalisation, transfer, and re-verbalisation, whereas others consider the concept more in more detail. Neubert (2000) suggested that there are seven features of translation competence: complexity, heterogeneity, approximation, open-endedness, creativity, situationality, and historicity, and there are five parameters of translational competence: language competence, textual competence, subject competence, cultural competence, and transfer competence (Neubert 2000: 6). Orozco (2000) suggested that the central element constituting translation competence is transfer competence, which can be further sub-categorised into comprehension competence, de-verbalisation competence, re-expression competence, and competence in carrying out the translation project. The other components of translation competence are communicative competence, extra-linguistic competence, instrumental-professional competence, psycho-physiological competence, and strategic competence (Orozco 2000: 199-201). Furthermore, PACTE (2003) defined translation competence as the translators' ability to carry out the transformation from the ST to the TT, while keeping the function of the original text and the receptors' needs in mind. The PACTE model of translation competence includes six sub-competencies specifying bilingual, extra-linguistic, situational, instrumental, strategic, and psycho-physiological knowledge. Developing translation competence, on the one hand, consists of the acquisition of the individual sub-competencies. It requires re-structuring of the sub-competencies acquired.

Besides acquiring new knowledge, skills, and techniques, two qualitative changes that take place in all kinds of the learning process are the restructuring of the learners' knowledge and fostering the ability to apply their knowledge to solve actual problems (Presas 2000: 29). According to Göpferich and Jääskeläinen (2009), with increasing translation expertise and experience, the translator is able to handle larger translation units (e.g., Gerloff 1988; Krings 1988; Jääskeläinen 1999), tackle more complex problems (e.g., Jääskeläinen 1999), show greater awareness of the target audience and translation problems, have a higher level of self-monitoring and revising, and use references more frequently (e.g., Gerloff 1988).

Interpreting competence and STR competence are also topics worth exploring. For instance, Gile (2009) added a coordinating effort in his *Effort Model* to emphasise the idea that interpreting competence also consists of the ability to coordinate ST comprehension and TT production. Meanwhile, the STR competence could be described as, but not limited to, the combination and interaction between reading and speaking in two different languages. It could be argued that it is similar to both written translation competence and interpreting competence because it shares either the input channel or the output channel with the two modalities. However, it is different from them due to its own characteristics.

To start with, STR requires well-developed reading skills (Nilsen and Monsrud 2015: 11). Based on a study of public-sector interpreters' reading speeds in Norwegian, Nilsen and Monsrud (2015) discovered that 70% of the subjects did not possess sufficient skills in *decoding*, which is the central component of reading (Nilsen and Monsrud 2015: 10). Nilsen and Monsrud (2015) argued that only by possessing highly automated reading skills, can interpreting trainees perform STR without being hindered by decoding problems. Hence, they suggested interpreter training programmes to include not only reading training in the curriculum but also a reading test in the admission procedure (ibid.: 17).

A successful STR production should sound as if the document was written in the target language, meaning that it demands a smooth oral delivery without redundancy, hesitations, and inappropriate pauses (Angelelli 1999: 27). STR requires, however, more than verbal skills. Weber (1990: 50) summarised the specific skills required for performing STR as: the ability to analyse the text rapidly, to transfer information from one language to another immediately, to avoid literal translation, and to deliver the TT smoothly. Furthermore, Stansfield (2008: 10) described the ideal sight translator as someone who is highly literate, highly proficient in both the source language and target language, possesses rapid interpretation skills, has a solid knowledge of the subjects, and is familiar with rules, expectations, and requirements of the clients and parties that solicit the services.

The *expert-novice paradigm* (Timarová 2010: 134) has been employed in many written translation and oral interpreting studies (e.g., Hoffman 1997;

Moser-Mercer *et al*. 2000; Alves, Pagano, and Silva 2009; Bayer-Hohenwarter 2010; Prassl 2010) to uncover the relationship between translation competence and many other variables. These studies have yielded many interesting results. For instance, Göpferich (2010) compared twelve students with ten professional translators and proved that the professionals made fewer errors than novices. Bayer-Hohenwarter (2010) recruited four first-semester students, three third-semester students, and five professionals to compare these three groups' behavioural differences during translation. She found that "successful translators strike a cognitively efficient balance between flexible problem solving and routinized reflex" (ibid.: 83).

Comparing the performance of translation trainees and professionals is another way of investigating STR competence. In a paper presenting the preliminary findings of a pilot study examining six student interpreters' and three professional interpreters' STR accuracy and quality, Lee (2012) suggested that less experienced interpreters need to make a conscious effort to develop their reading skills further to avoid literal translations. In comparison, student interpreters made more major accuracy errors, and the issues in their target language revealed that they had greater difficulty in coordinating the reading process and the oral production. Based on the analysis of the qualitative differences in these two groups' performance, the researcher argued that "the condensation strategy can serve as an important skill enabling the interpreter to shorten the delivery time and avoid wordy, often awkward, target language expressions" (ibid.: 710). Furthermore, Lee (2012) called for more research to help develop effective pedagogical methods to enhance student interpreters' ST performance. However, the task of recruiting participants is sometimes problematic. Assuming that advanced interpreting trainees would display more efficient reading patterns during STR than less experienced trainees, Chmiel and Mazur (2013) carried out an eye-tracking study, but they found no significant evidence to corroborate their hypotheses. This finding could be due to the one-year difference in training between the two groups not being sufficient enough to result in greater reading efficiency (ibid.: 1).

The relationship between eye movements and reading proficiency has attracted some attention from researchers. Carver (1982, 1992) and De Luca *et al.* (2013) shared the idea that after years of reading practice, readers can develop suitable idiosyncratic styles and an optimal rate of reading. Reading skill level can influence the characteristics of readers' eye movements. Less skilled readers have longer fixations, shorter saccades, and more regressions than skilled readers (e.g., Rayner 1998, Ashby *et al.* 2005). Therefore, inexperienced and experienced readers will display distinctive eye movement patterns during reading aloud (RA). Moser-Mercer (1995) argued that during STR, beginner interpreters are inclined to attempt a semantic and referential interpretation when they encounter linguistic units in the ST, especially difficult ones. On the contrary, experienced interpreters tend to adopt a non-linear, meaning-driven approach. However, the generalisation that beginner interpreters' reading pattern is straightforward and linear is oversimplified as there are many more intricate cognition-related activities taking place during STR practices (Zheng and Zhou 2018). Nevertheless, Moser-Mercer (1995) did emphasise the difference between the behaviours of beginners and experienced interpreters during STR, leaving us the opportunity to further explore the characteristics of the differences and the reasons behind them.

2.4 STR Directionality

Translation directionality is a common topic in translation research. Research on translation directionality discusses how translators/interpreters translate their first language into a foreign language, and from a foreign language into their first language. Of course, it does not exclude translating/interpreting between two non-L1 languages (Pavlović 2007: 3). In the context of the present research, I hope to find indicators that can be applied to address STR directionality. Contrarily, it is also hoped that studying STR directionality can instigate and examine new indicators.

According to Pedersen (2000: 109), the notion of first language (L1) during translation does not necessarily mean it is the first-acquired language, but "the language that is most readily available" to the translator. Second language (L2) is another language the translators/interpreters have mastered to a high level of competence, although it is not necessarily the second-acquired language in the translators'/interpreters' lifetime (Pavlović 2007: 81). Translating from one's L1 to L2 is termed *forward* or *reverse* translation;

translating from one's L2 to L1 is called *backward* or *direct* translation. Macizo and Bajo (2004: 199) have pointed out that theorists, such as Kroll and Stewart (1994), have proposed "an asymmetrical relation between the two represented languages L1 and L2 in the bilingual's mind". On a textual level, Macizo and Bajo (2004) argued that "[T]he lexical connections between L2 and L1 are stronger than those between L1 and L2. In contrast, conceptual connections are stronger in forward translation (L1 to L2) than in backward translation (L2 to L1)" (2004: 199). In other words, translators focus on conceptual and functional information in the ST during L1-L2 STR because of their superior L1 reading competence and a better grasp of superficial lexical meanings in the ST during L2-L1 STR because of their relatively lower L2 reading competence. To conclude, forward translation and backward translation are inevitably different because the language-processing mechanism involved and activated during the two tasks is not the same.

Newmark (1988: 3) argued that "translat(ing) into your language of habitual use […] is the only way you can translate naturally, accurately and with maximum effectiveness". This idea might have started the discrediting of L1-L2 translation. Hence, traditionally, translating from one's native language into a foreign language was nothing more than a pedagogical means to practice certain grammatical features (Pavlović 2014: 150). Nowadays, many translation organisations still believe that "ideally all translations must be done by native speakers of the language of the target culture" (Grosman 2000: 17). Therefore, it is suggested that translation directionality should be used as "an operational parameter to separate routine from non-routine tasks" (Saldanha and O'Brien 2014: 116). According to Pavlović (2014: 150), this view is "still widely present in Europe, which can be supported by the fact that international organizations only accept the translation into the mother tongue". As a result, the fact that "translation into a non-mother tongue was completely rejected and disapproved [of]" was thought to be the reason for the lack of theories and research interests on L1-L2 translation (ibid.: 149).

However, in the past two decades, the number of studies investigating topics related to L1-L2 translation has increased (Pavlović 2007: iv). In 1998, Campbell (1998: 4) had already challenged the traditional view by describing L1-L2 translation as "an activity as normal and possibly as widespread as translation into the first language". Later, researchers such as Pokorn (2005: 37) criticised the traditional view of "ignoring the practice of L2 translation and accepting the assumption that translators should work only into L1". Questioning Marmaridou's (1996: 60) claim that "translating into one's mother tongue generally yields better texts than translating out of it", Pavlović (2014) conducted research on the quality of translation from the translators' L2 into L1, and from their L1 into L2, in the hope of determining the characteristics and the relationship between forward and backward translation. As mentioned before, the traditional standpoint tended to ignore the importance of L1-L2 (forward) translation. The results of Pavlović's (2014: 149) study, however, indicated that both ways of translating "are possible and that both directions include certain difficulties", and backward translation "is not free from flaws and that it does not come naturally as many are determined to claim" (ibid.: 162). Hence, Pavlović (2007: 81) disagreed with the traditional view regarding the direction of translation and emphasised that directionality should not be an excuse for low-quality forward translations in the workplace.

Some researchers, such as Campbell (1998: 57), argued that "translating into a second language is very different from translating into the first language". However, in the 21st century, researchers became interested in learning the similarities between the two forms of translation. For example, Pedersen (2000: 110) argued that both translation scenarios contain two difficulties: understanding the implications of the ST and rendering them into the TT adequately. Pavlović's (2007) study acknowledged both differences and similarities between the two translation processes. One important conclusion was that the novice translators in the study "tend to encounter similar problems, and respond to them with a similar blend of actions/interactions, regardless of direction of translation" (ibid.: 187).

Using eye-tracking, Pavlović and Jensen (2009) carried out a study on translation directionality in which student and professional translators were asked to translate two comparable texts from their L1 (Danish) into their L2 (English) and the other way around. Their study and the present study share some similarities in the choice of research methodology and subjects. In addition, they both touch upon STR directionality. By looking at four indicators of cognitive effort, the gaze time, the average fixation duration, the total task length, and the pupil dilation, Pavlović and Jensen (2009) concluded that TT processing demands more cognitive effort than ST processing in both forward and backward translation. The hypothesis that L1-L2 translation tasks require more cognitive effort than L2-L1 translation tasks, however, was only partially confirmed by an increase in task duration and pupil dilation (Pavlović and Jensen 2009: 107). Without any supporting evidence from the total gaze time and mean fixation duration, the study, therefore, suggested that L1-L2 translation "may *not* necessarily be more difficult than translation into L1, as is widely assumed" (ibid.). In fact, according to Pavlović's (2007: 169) introspective interview data, inexperienced student translators subjectively found L1-L2 translation easier than L2-L1 translation. However, Pavlović and Jensen (2009: 108) acknowledged that it would be premature to draw any definitive conclusions due to the limited amount of data collected and the incomparability of the STs.

Pavlović and Jensen's (2009) eye-tracking study revealed two significant difficulties in conducting translation directionality studies. Firstly, they found it extremely difficult to "explain the discrepancies between the various indicators of cognitive effort when it comes to L1 and L2 tasks on the whole" (2009: 107), reminding other researchers that the choice of eye-tracking data as indicators is also of great importance. Secondly, their study indicated that the comparability of texts is always a problem when it comes to comparing either the processes or the products of forward and backward translation. Nevertheless, their findings are very inspiring as they have challenged traditional assumptions about translation directionality. Pavlović and Jensen's (2009) research informs other researchers that when two STs are not comparable, researchers can only draw a conclusion relevant to the specific experiment, even if the eye-tracking data collected in the two translation processes showed significant differences. However, the researchers could conclude that one of the tasks is more cognitively demanding than the other in that study. To be specific, a forward translation task might require more effort than a backward translation task in one study, but the researcher cannot generalise and conclude that all forward translation tasks require more effort than backward translation tasks.

2.5 Reading Aloud (RA) and STR: Different Reading Patterns

Researchers can easily associate STR with written translation and oral interpreting because they are all translation activities. However, STR is strongly related to a reading modality due to their similar inputs and outputs. This reading modality is Reading Aloud (RA). Compared to STR, RA has been used and studied extensively. Some might wonder what the point of studying mature human subjects' RA process is, since RA is mostly used in the preliminary educational scenario. From the perspective of the present research, if researchers only focus on the preliminary educational scenario and refuse to know more about RA processes at an advanced level and what RA competence consists of, how can they improve their knowledge of RA? The present research, therefore, hopes to improve our understanding of both reading-speaking activities, namely STR and RA, by comparing subjects' cognitive processing in experiments.

The notion of *reading* is broad (McConkie 1983: 65). There are different reading modes, for example silent reading and reading aloud (RA). There are reading modes with different purposes such as reading in order to recite, reading in order to translate, and reading in order to answer questions. Long before researchers started to devote attention to the reading process during translation, there have been numerous studies of readers' eye movements (e.g., Just and Carpenter 1980; Rayner 1998). In the mid-20th century, researchers already came to realise that the nature of the reading tasks and the materials determine the eye movement patterns (e.g., Ledbetter 1947; Tinker 1951). When those researchers strived to investigate the reading process, their subjects' eye movements provided valuable information (Gibson and Levin 1975: 351). When Rayner (1998: 373) was trying to gauge fixation duration in various visual and reading tasks, the different results revealed that "the nature of the task codetermines the duration of the fixation", and, therefore, he reminded us that "it is important to interpret the fixation data in light of the *kind* of reading that the translator is performing" (ibid.). Researchers have to acknowledge the differences in the perceptual processes that are involved in these two different reading-related tasks, but they should be careful not to overgeneralise (McConkie 1983: 65). However, this does not mean researchers cannot compare different reading processes.

RA is an example of mapping visual input to a verbal representation (Timarová 2010: 137). Also called oral reading, it has been compared with silent reading in many different ways. It is viewed as “the default in classical antiquity” that “processes silent reading in individual development, for example in primary school education” (Laubrock and Kliegl 2015: 2). Before it is replaced with silent reading by competent readers, RA is a common strategy for beginner readers (Krieber *et al*. 2017: 2). In silent reading, readers’ short-term memory is filled with the meaningful segments, instead of individual letters (Levin and Addis 1979: 32). Although “global understanding was not affected by the type of reading” (Macizo and Bajo 2004: 198), RA differs from silent reading in many ways because the reader has to distribute his/her attention to each of the words. In the earlier phase, it has been found that the readers had better results in the memory tests after silent reading than after oral reading. For example, in a research project by Buswell (1927), the subjects reported that instead of grasping the meaning of the text, they were paying more attention to the pronunciation of the words during oral reading. Nevertheless, later researchers suggested that besides the verbalisation aspect involved in the RA process, there is little or no difference between the central processes underlying the comprehension process in the two types of reading (Levin and Addis 1979: 37; Krieber *et al*. 2017: 2).

Researchers, such as Krieber *et al*. (2017), have discovered that the reading mode significantly influences both the spatial and the temporal characteristics of readers’ eye movement patterns. Although “the eye movements in both oral and silent reading are largely controlled by the recognition of the meaning” (Buswell 1920: 99), it has been found that readers have a higher fixation count (e.g., O’Brien 1926; Wanat 1971) in RA than in silent reading. Regarding the cognitive workload of different reading modalities, it was found that “oral reading requires more visual attention than the silent reading of the same type of materials” (Wanat 1976: 133). This would imply that RA is more cognitively intense due to the extra attention readers allocate to each word (Buswell 1937).

Reading purposes and reading materials influence not only the subjects’ gaze behaviour when reading, but also their eye movements in translation activities (Jakobsen and Jensen 2008: 120). Schaeffer *et al*. (2016: 208) stated that “a multitude of concurrent processes are at play during (reading for) translation”. Therefore, it is no surprise that comprehension is slower during reading for translation than during regular reading. Also, the comparison between reading for comprehension and reading for translation showed the differences in the investments of cognitive effort and the allocation of attention. It clearly indicated that the comprehension process varies with the specific goal of each reading task (Macizo and Bajo 2004: 181).

Reading for RA and reading for STR are two similar but different reading processes. Both the RA and the STR reading processes consist of word identification and word interpretation. *Word identification* refers to “the singling out of an element within the lexicon” (Ehrlich 1983b: 193), whereas *word interpretation* refers to “the establishment of the relationship between the word on the page and the other concepts that have appeared earlier in the text” (Ehrlich 1983b: 194). It is the Collaborative Activation-based Production System (CAPS) that allows this variety of processes to occur concurrently during reading (Carpenter and Just 1983: 285). Compared to silent reading, RA and STR are both “integrated series of cognitive processes that operate at different levels of abstraction and that occur simultaneously” (Wolverton and Zola 1983: 41), but they are much more complicated than silent reading in terms of how many levels of abstraction are involved.

The comprehension processes of reading for comprehension and reading for translation both involve speech processing, lexical access, sentential processing, and discourse processing (Macizo and Bajo 2004: 181). In the same vein, the comprehension processes of RA and STR include the same set of procedures. To start with, the need to produce accurate and fluent speech at an appropriate speed exists in both RA and STR. As the researchers delve into the underlying process of RA and STR, they pointed out that “word comprehension involves memory search of the abstract, internal lexicon and that there are parallel coding processes involving multiple neural pathways” during both RA and STR (Downing and Leong 1982: 191). The subjects see the ST, decode the meaning, store it in their short-term memory, arrange their thoughts, and then verbalise the content, either in the same language or in another language. Both types of reading require a division of the readers’ attention between their reading input and speaking output. However, since the linguistic factors affecting the reader's allocation of visual attention may be different in reading aloud and silent reading (Wanat 1976: 133), they could be different in oral reading and STR as well. The present research explores this further in the description of the experiment and reports in Chapter 6.

De Luca *et al*. (2013: 1) emphasised the complexity of RA, saying that it is "a complex task that requires the synchronization of various subtasks or sub-components which impinge on different ongoing fluxes of information". STR, on the other hand, requires the synchronisation of even more subtasks. Compared to RA, STR involves an additional set of cognitive operations: a code-switching process between the two languages (Macizo and Bajo 2004: 182). From a *vertical perspective* of translation, Seleskovitch (1976) placed three processes in a sequential order: comprehension, code-switching, TT production. From a *horizontal view*, however, the three sets of operations are taking place at the same time. For example, Macizo and Bajo (2004) believed that a code-switching process has already taken place before the ST comprehension has been completed. Meanwhile, the STR process involves more pathways and solutions. Hence, although the fundamental properties of the eye movements in STR are similar to those in RA, the trigger mechanism involved in deciding when and where to move is different.

During RA and STR, three types of memory mechanisms are involved. The first one is the *iconic memory*. It is "a temporary memory store in which much of the information physically available in the stimuli is still available after the display has gone off" (Rayner and Pollatsek 1989: 15). Although iconic memory is highly transient, it has a large capacity (Rayner and Pollatsek 1989: 17). During both RA and STR, the physical information of the reading material/the ST would have been stored in the iconic memory. Because of the transient nature of iconic memory, relatively important information is registered in a more permanent structure called *short-term memory* (ibid.). Sometimes termed as *working memory* (Baddeley 1986), it is "a flexible workplace whose limited capacity can be allocated to either storage or processing" (Rayner and Pollatsek 1989: 18), and information registered in it "can remain as long as it is being worked on" (ibid.). Despite its small capacity, the working memory is a crucial cognitive component necessary in the translator's overall performance (Timarová 2010: 137).

One’s short-term memory is highly activated in STR due to the almost simultaneous nature of this translation mode. During STR, chunking allows the translators to grasp the meanings of the segments and to store the meanings in their short-term memory when orally translating them into another language. Such a temporary storage unit in the information process is sometimes referred to as one’s *buffer* (Rayner and Pollatsek 1989: 17). The third type of memory mechanism is the *long-term memory*. Of course, long-term memory plays a certain role during both RA and STR, and it can be triggered by certain information. Nevertheless, RA requires a faithful oral copy of the text and involves less problem-solving while STR demands translators to seek more evidence and support from their long-term memory to solve translation problems. Therefore, from the perspective of the present study, long-term memory is perhaps more activated in STR than in RA.

In the next chapter, the present study introduces and analyses a phenomenon that is relevant for the analysis of both RA and STR: The Eye-Voice Span (EVS).

Chapter 3. Theoretical Framework

3.1 Cognitive Psychology and STR Research

Cognitive psychologists became very interested in studying reading processes in the early 1970s (Rayner and Pollatsek 1989: 3). Their primary methodology was empirical experimentation (Rayner and Pollatsek 1989: 8), which was later adopted by researchers in the field of translation studies. Gaining popularity in 21st century, Translation Process Research (TPR) became "a branch of descriptive translation studies that investigates the underlying cognitive and mental processes rather than the products resulting from human translation" (Läubli and Germann 2016: 159). Its primary aim was to understand translation processes through observation. Translators' and interpreters' cognitive resources and attention allocation is one of the most popular topics in the TPR. Cognitive resources are the mental capacity used for processing information, while attention allocation is the activity of allocating cognitive resources to completing a specific task (Hvelplund 2011: 38-9).

Many process-oriented translation studies yielded interesting results. For example, Hvelplund's (2011) process-oriented research on translators' allocation of cognitive resources during written translation revealed the following: firstly, processing the TT is more cognitively demanding than processing the ST. Secondly, professionals and less experienced translators allocate a different amount of cognitive resources to the same task. Thirdly, STs that are more complex required more cognitive resources compared to less complex texts. Finally, translators' attention allocation responds differently to time pressure. There are many other topics when researchers are investigating translators' allocation of cognitive resources during STR. After all, STR is different from written translation in many ways. For example, compared with written translation, STR has a more or less stringent time limit depending on the task demand, and it allows translators less opportunity for self-correcting or revising. For example, in some of the oral interpreting tasks, a division of attention seems to be possible during STR. In Sampaio's (2007: 66) words, STR is as "a *multi-task* that requires close intense concentration, specific skills and accuracy". However, although "simultaneous interpretation is a classic case of *divided attention* in that it involves several different cognitive tasks carried out more or less concurrently” (Lambert 2004: 297), it cannot be applied directly to other translation modalities such as STR, which involves oral outputs from visual inputs.

One theory regarding the limits of human attention is that the brain acts as a single communication channel of limited capacity in its moment-to-moment decision-making (Craik 1948; Welford 1952; Broadbent 1958). This is called the single channel hypothesis (Deutsch and Deutsch 1963; Neisser 1967; Norman 1968), which, however, was disproved by Allport, Antonis, and Reynold’s (1972) study. In fact, several studies have required subjects to perform two tasks simultaneously (Welford 1968; Allport *et al.* 1972). In Spelke, Hirst and Neisser’s (1976) study, participants were able to achieve a division of attention in extracting meaning simultaneously from what they read and what they heard after several weeks of practice. Although the result was achieved through training before the experiment, it provided evidence against the hypothesis that human beings have a limited attention capacity and only one cognitive channel (Spelke, Hirst and Neisser 1976: 98). Researchers argued that allocating attention to two tasks simultaneously was possible because at least one of the two actions was being carried out automatically without conscious control (Hirst *et al.* 1980; Solomons and Stein 1896); the attention alternated rapidly between the two activities (Jaffe *et al.* 1967); or a genuine division of attention was accomplished (Downey and Anderson 1915). These ideas became a series of hypotheses concerning attention division between synchronous tasks: the extra-effort hypothesis, the alternation-of-attention hypothesis, and the automatic-mental-activities hypothesis (Lambert 2004: 298). If we look at this series of hypotheses together with Gile’s (2009) *Effort Model* for STR (STR = Reading + Short-term Memory + Production + Coordination), it is possible to suggest the following: a) managing reading input and oral output demands extra effort than carry out one of the two activities; b) coordinating reading and speaking involves attention shifts; c) devoting cognitive effort to short-term memory and coordination is, to some extent, an automatic mental activity.

In addition, the division of attention has been found to be possible in written translation. Based on the mechanical pause-based segmentation, Dragsted and Hansen (2008) found that a pause-defined segment usually involves comprehension, production, and coordination. On average, mixed ST and TT segments that involve both reading and typing constitute 58% of the segments in total, meaning that the subjects' attention was divided between both ST and TT processing at the same time in more than half of the cases. Sometimes reading the ST and producing the TT were corresponding counterparts of the same phrase in two languages, but sometimes they were not.

Previously, Dejean Le Féal (1981) compared *sight translation* with *simultaneous interpreting with text*. He concluded that it is impossible to process a visual input and an audio input at the same time. Whether it is possible to process a visual input and produce an oral output at the same time, however, remained almost unexplored until Huang (2011) carried out a study investigating a reading behaviour called *reading ahead*. Reading ahead refers to the forward reading that happens when an interpreter's oral output is lagging behind. Huang (2011) found that *reading ahead* occurred at most of the sentence boundaries in her experiment, indicating an overlap between reading and oral production during Chinese-to-English STR (2011: 64). Furthermore, STR shares some similarities with another non-translation task concerning the formats of inputs and outputs: reading aloud (RA), which is also known as oral reading. With reading as its input and speaking as its output, RA requires attention allocated to each word (Buswell 1937), and demands a division of the readers' attention between their eyes and voice. Therefore, the present study suggests that: on the surface, the brain needs to have some visual input to work on before it can start to instruct the speech organs to produce oral output causes EVS, which is the main focus of the next section; and below the surface, such performance are possible due to attention division on sub-efforts: reading, short-term memory, production and coordination (Gile's *Effort Model* for STR 2009).

3.2 Understanding Eye-Voice Span (EVS)

Having a basic understanding of both spatial and temporal Eye-Voice Span (EVS) is extremely important before one can move on to investigate them in empirical studies. This section starts with a brief introduction of vision and perceptual span, which is a concept that might be confused with EVS. It then defines spatial and temporal EVS, reviews the measurement of both the spatial distance and temporal latency, and looks at the application of temporal EVS in eye-tracking based STR studies.

3.2.1 Vision and Perceptual Span

When humans look at things, images are turned upside down in the lens and then projected onto the retina. The light-sensitive cells on the retina, called cones and rods, are receptors responsible for transducing "the incoming light into electrical signals sent through the optic nerve to the visual cortex for further processing" (Holmqvist *et al*. 2011: 21). The cones are extremely well represented in a small area named fovea but are sparsely distributed elsewhere on the retina. The density of cones is inversely proportional to the distance from the fovea (Rayner and Pollatsek 1989: 9). As a result, one only has "full acuity" in the fovea (Holmqvist *et al*. 2011: 21). To see a word in a text, one must manage eye movements to let the light from that word fall on the fovea, and the foveated information is prioritised in processing (ibid.).

In general, the central fovea area covers about two degrees of the total vision field (Downing and Leong 1982: 137; Rayner and Pollatsek 1989: 9). According to Rayner and Pollatsek (1989: 9), "a horizontal line of text falling on the retina can be divided into three regions: foveal, parafoveal, and peripheral". While acuity is greatest in the fovea area, it "drops off markedly in the parafovea and even more so in the periphery" (ibid.). The parafoveal area "subtends about 10 degrees of visual angle around fixation (4 degrees to the left and to the right beyond the foveal region)" (ibid.: 9). The peripheral vision, which covers "everything on the line of text beyond the parafoveal region" (ibid.), is "not nearly as acute as that in the central foveal region" (Downing and Leong 1982: 137). Some researchers considered it difficult to judge how useful peripheral vision is because readers are only able to "detect little useful information besides interword spaces and the lengths and shapes of words" (Gibson and Levin 1975: 356). Nevertheless, it was proposed by Shebilske (1975) that the function of the peripheral vision is predicting the next position to fix one's eyes on and guiding the eye movements accordingly. In this way, the peripheral vision aids the reading process, and it enables the reader to move his/her eyes across larger units, rather than from one word to the next.

Perceptual span is sometimes called visual span or span of apprehension (Downing and Leong 1982: 143). McConkie (1983: 81) defined it as “that region around the centre of vision within which some aspect of visual detail of interest is used in reading (or affects the reading process)”. Keating (2013: 72) has simplified the definition as “the amount of useful information that a reader can extract from a text on a given fixation”. The moving window paradigm, for example, was often applied to investigate this topic. The reader’s perceptual span helps to identify the terminal point of the current meaning unit and the initial point of the next. Hence, the essence of the perceptual span is that it can tell us how much information can be perceived during a fixation (Rayner and Pollatsek 1989: 5). With a fixation as its centre, the perceptual span is usually asymmetric. For readers of languages that are read from left to right, the perceptual span on the right side is wider than that on the left side of their fixations. Contrarily, for readers of languages that are read from right to left, the perceptual span on the right side is narrower than that on the left side of their fixations (Keating 2013: 72-3).

Measurement of the average perceptual span is rather straightforward: it is found by dividing the total number of words in the passage by the number of fixations of the eyes (Downing and Leong 1982: 143). The length of one’s perceptual span, however, might change with the task and the text (McConkie 1983: 81). Based on various studies he and his colleagues conducted (e.g., McConkie and Rayner 1975; Rayner, Well, and Pollatsek 1980), Rayner (2009) concluded that readers of alphabetic languages can obtain information from approximately 3-4 characters’ spaces left of a fixation and 14-15 character spaces to the right of the fixation (2009: 1462). According to Keating (2013: 72), however, readers usually cannot identify words that appear beyond seven to eight characters to the right of a fixation due to the anatomical limitation in visual acuity. The difference between these findings might be the result of varying research settings, subjects, or experimental designs. Its length also depends on the language of the written text and the nature of the reading task. The perceptual span for Japanese, a logographic writing system, is considerably smaller than that for English (Rayner and Pollatsek 1989: 134). As another character-based written language, Chinese generates a narrower perceptual span when read by its native users. It is about one character to the left and two to three characters to the right of a fixation (Keating 2013: 73). Hence, it is possible that a native Chinese reader may appear to have a narrower perceptual span than a native English reader when they are both reading a text written in English. Moreover, it could also be the case that the size of the perceptual span is dependent on the reader's language competence (Rayner 1986; Rayner 2009). Furthermore, compared to silent reading, RA "may use more attentional resources and lead to higher foveal load than silent reading, resulting in a reduction in the perceptual span" because of "the additional demands of articulation and the associated scheduling and coordination" (Pan *et al*. 2017: 261). However, Perceptual span should not be confused with Eye-Voice Span (EVS), which is discussed in detail in the following sections.

3.2.2 Defining Spatial and Temporal EVS

Some phenomena similar to EVS have been investigated in parallel with reading/performing skills, such as music and typewriting. For example, there is the eye-performance span in playing a musical instrument or singing (Thompson 1987; Fine *et al*. 2006); the eye-audio span in music literacy, including reading vocal and instrumental music (Jacobsen 1941; Silva and Castro 2018); the eye-hand or eye-finger span in typewriting (Book 1908; Shaffer and Hardwick 1969). Later, researchers in the field of translation and interpreting discovered similar parameters and applied them in translation process research. For instance, the ear-voice span, which describes the span between the input and the output during re-speaking (Chmiel *et al*. 2017) and simultaneous interpreting (e.g., Lee 2002; Christoffels and De Groot 2009); and the eye-key span in written/typed translation (Dragsted and Hansen 2008; Dragsted 2010; Carl and Schaeffer 2016). All these measurements have both a spatial and temporal element. Eye-key span, for instance, can be calculated by words, letters, or milliseconds. Some researchers used the term *EVS* to address another type of span: the *ear-voice span* in simultaneous interpreting (Dragsted and Hansen 2008: 21). The term EVS in the present study, however, only refers to Eye-Voice Span.

The first reference to spatial EVS is found in a study by Quantz (1897), whose focus of study was the psychology of reading. A decade later, Huey

(1908) considered EVS when researching psychological and pedagogical aspects of reading. Since 1920, Buswell has given spatial EVS in reading a profuse amount of attention and substantial coverage in his studies. Buswell (1920) defined spatial EVS as the spatial distance that the eyes are ahead of the voice during RA. Later, Morton (1964: 347) refined the definition of EVS as “a measure of the amount of material or time by which the voice lags behind the eyes in oral reading”. It is usually measured by “time, letters, letter spaces, ems (a printer’s measure), syllables, or words, which is the most common index” (Levin and Addis 1979: 1). From the definition made by Morton (1964) and the calculation method used by Levin and Addis (1979), one can see that researchers did not put much emphasis on distinguishing spatial EVS and temporal EVS. However, spatial EVS and temporal EVS should not be conflated. The present research, therefore, hopes to make a clear distinction between these two measurements.

Spatial EVS (De Luca *et al.* 2013), is the material span sometimes referred to as *Eye-Voice Distance* (Inhoff *et al.* 2011). The two ends of a Spatial EVS fall on what the eyes are fixated on and what is being articulated at the time. Levin and Addis (1979: 47) agreed that “an eye-voice span of considerable width is necessary in order that the reader may have an intelligent grasp of the material read, and that he may read it with good expression”. According to Levin and Addis (1979: 1), spatial EVS already had “a long and useful history in the annals of educational and psychological research” in the last century. The study of spatial EVS also has a long history in reading research (Downing and Leong 1982: 145). It appears that initially, spatial EVS has been studied as an interesting phenomenon in and of itself, but researchers later started to focus on the relationship between spatial EVS and the readers’ reading skills. Since the 1960s, psychologists’ and educators’ main interest have shifted to understanding the process of reading through studying spatial EVS, meaning that analysing spatial EVS has become a research method (Levin and Addis 1979: 2).

Temporal EVS, also called Fixation Speech Interval (FSI) (Inhoff *et al.* 2011), or Eye-Voice Latency (EVL) (Holmqvist *et al.* 2011), is the temporal latency between the time that a certain point, usually a phoneme, has been read by the eyes and the time that it was articulated. Gibson and Levin (1975:

360) stated that having a temporal EVS is natural and inevitable during RA because “in order to read with normal intonation the reader must have information about the sentence which occurs to the right of the word he is actually reading aloud”. Some have argued that a readers’ eyes could “pick up cues about upcoming words in peripheral vision” (Levin and Addis 1979: 3) and thus add to the duration of temporal EVS. When investigating the impact of the perceptual span on the eye movements of reading a text written from left to right, researchers found that if a word (W_1) is situated on the right side of a previously fixated word (W_0), W_1 might have been visible during the previous fixation. W_1, therefore, is likely to be read faster visually (Rayner 1998) or skipped (Rayner 2009). However, it has been argued that the peripheral cues are not within the clear vision and so they are not sufficiently adequate to influence the duration of temporal EVS (McConkie and Rayner 1975). Even if they are, the preview benefit is too slight to be influential, as the *parafoveal preview benefit* is reported to be only 30-50 ms typically (Keating 2013: 74).

3.2.3 Measuring EVS during RA

The spatial distance of spatial EVS and the temporal duration of temporal EVS are “tightly coupled aspects of relative motion” between visual and vocal events (Holmqvist *et al*. 2011: 428). However, spatial EVS and temporal EVS should not be used interchangeably as spatial EVS is in the spatial domain, whereas temporal EVS is measured in the temporal domain. Below is a review of studies that measured either spatial EVS or temporal EVS with the help of different apparatus.

Many researchers, such as Buswell (1921) and De Luca *et al*. (2013), have measured the average spatial size of spatial EVS. It was found that the size of spatial EVS in RA changes as the eye moves in advance of the voice, sometimes relatively far and sometimes not very far ahead (Buswell 1921: 217). When Quantz first published about spatial EVS (1897), the simplest way to observe and measure spatial EVS during RA was to make the text unavailable. By covering or removing the text when a reader read to a predetermined point, researchers could record how many words the reader had articulated beyond the critical position. Quantz’ (1897) spatial EVS was calculated by measuring the number of words produced within the gap between reading offsets and articulation offsets. The number of words was then accounted as a spatial distance. To be specific, if a subject articulated three words after the reading material went out of sight, the spatial EVS was counted as three words in this case. In a similar manner, Huey (1898) measured the length of spatial EVS during RA by noting down how many syllables on the previous line have been uttered after the subject had started the sweep to the beginning of the next line. Huey's (1898) spatial EVS was therefore calculated by counting syllables articulated within the gap between saccade onsets and articulation offsets. In another study by Huey (1908), spatial EVS was measured when the subjects turned the page, by counting spatial distance formed by the content articulated between the time of their reading offsets and their articulation offsets.

Later, Gray (1917) noted down oral output produced between the time of fixation onsets and articulation onsets to calculate spatial EVS, using an apparatus that supervised the reading process by projecting and monitoring a beam of light that could reflect the location of the eye movements precisely. Buswell (1921) carried out the calculation by contrasting the onsets of eye movements and the onsets of subsequent verbal articulation, using an apparatus that photographed a beam of light reflected from silvered glass mirrors to the cornea of the eye and then through a lens to a moving film. However, for those who recorded the subjects' eye movements photographically, it was difficult to keep the subjects' heads immobile. Various types of chin rests and head clamps were trialled, but none were found to be entirely satisfactory (Levin and Addis 1979: 14).

Due to technical problems and concerns over data quality, researches took a step back and started to apply the simplest method again. Researchers such as Lawson (1961), Levin and Turner (1968), Bond and Tinker (1973) chose to calculate spatial EVS by setting a light-out position in the reading material. When a subject read to a predetermined position, the researcher would switch off the lights or the screen and count the number of words that the reader could produce orally. Using this method, Stuart-Hamilton and Rabbitt (1997) estimated spatial EVS during RA to be approximately 4 words; Levin and Addis (1979) suggested that the average readers' usual spatial EVS ranged from 3 to 5 words, whereas skilled readers could have a spatial EVS as large as 8 words that reduces when necessary.

In the last century, spatial EVS in RA has been traditionally measured using the off-line methods mentioned above. Nevertheless, irrespective of whether the reading materials were removed or the lights were switched off, the reliability of the measurement was always questioned (Levin and Addis 1979: 16). In the 21st century, spatial EVS is measured by some relatively modern apparatuses. For example, De Luca *et al*. (2013) used a head-supported eye link 1000 eye tracker to measure the average gaps between the gaze and voice of both dyslexic and normal readers. In this study, spatial EVS was calculated by the material span between when a syllable was articulated and when that point was viewed. They reported that the normal readers' spatial EVS was about 14 letters.

Measuring temporal EVS during RA started later than the investigation of spatial EVS. This was mostly due to technological and methodological limitations. However, researchers could also measure temporal EVS through the off-line methods mentioned above, but they chose not to. Perhaps researchers preferred to measure temporal EVS when they were able to track eye-movements in a more precise way. Morton (1964), for instance, has made remarkable methodological progress by using electro-oculography to measure eye movements, although the quality of the recording was affected by electronic noise. Morton (1964) recorded subjects' eye movements with the electro-oculography technique, which "utilizes the presence of a standing potential difference between the front and back of the eye-ball" to detect the position of the eyes (Morton 1964: 341). This method noted down the onsets of eye movements and voice. It was reported that the average duration of temporal EVS was around 240 ms and there was not much variation of this value between different passages or readers (ibid.: 340).

Temporal EVS was also measured in some recent studies by using the most modern apparatus. Inhoff *et al*. (2011) used a head-mounted Eye-link II eye-tracking system in their research. They decided to measure temporal EVS in the middle of sentences, by excluding the first four to six words and final words of sentences. The remaining words were called the critical words. By contrasting the onsets of fixations on these critical words with the onsets of voice articulation of these words, they estimated an average temporal EVS of 486 ms (2011: 548). Inhoff *et al.* (2011) also discovered that a very long temporal EVS "was often corrected with a regression that moved the eyes closer to an articulated word" (ibid.: 554). These findings have been tested in a very recent paper by Laubrock and Kliegl (2015), who presented the source sentences on a 22 inches Iiyama Vision Master Pro 514 CRT monitor controlled by a custom C++ programme running on a standard computer. Unlike Inhoff *et al.* (2011), Laubrock and Kliegl (2015) did not make any distinction between words at different positions in a sentence. They calculated temporal EVS with two different algorithms: onset-to-onset temporal EVS and offset-to-onset temporal EVS. Onset-to-onset temporal EVS is the time span from the beginning of the first fixation on a word to the onset of its pronunciation. Offset-to-onset temporal EVS is the time span from the offset of the last fixation on a word to the onset of its pronunciation. The average onset-to-onset temporal EVS was 561 ms; the average offset-to-onset temporal EVS was 254 ms (Laubrock and Kliegl 2015: 6). The reason that their average offset-to-onset temporal EVS was significantly shorter than their average onset-to-onset temporal EVS is that the onset of articulation is the same in the two algorithms, whereas the offset of the last fixation takes place later than the onset of the first fixation.

By comparing the results yielded by applying the report-based procedures with results obtained in a selection recent studies (e.g., Järvilehto *et al.* 2008; Laubrock and Bohn 2008; Laubrock *et al.* 2007) that employ relatively modern methodology such as eye-tracking, Inhoff *et al.* (2011: 555) concluded that the spatial EVS estimated by the report-based procedure were inflated because readers could have guessed the following words. Laubrock and Kliegl (2015: 2) also criticised the off-line methods, such as the light-out method, because they ignored the readers' parafoveal preview, the guessing effect, and some task-dependent strategies. As a result, spatial EVS measured by using off-line methods were "grossly overestimated" (Laubrock and Kliegl 2015: 2). Those who applied the on-line methods tended to agree that the average temporal EVS in RA is approximately 500 ms (Laubrock and Kliegl 2015: 2).

In summary, calculating spatial EVS began earlier than calculating temporal EVS. Spatial EVS was often measured using off-line methods, and its physical length was equal to the number of syllables/words a subject articulated within the gap between the offset of reading input and the offset of speaking output. In contrast, temporal EVS was more frequently measured using on-line methods. The duration of temporal EVS is often calculated by onset-to-onset or offset-to-offset algorithms, although other algorithms are also possible (e.g. Laubrock and Kliegl 2015). To conclude, this section presents two common algorithms. Figure 3.2a is an example of an onset-to-onset algorithm that measures a temporal EVS from a fixation onset to an articulation onset of a syllable/word in RA. Figure 3.2b is an example of an offset-to-offset algorithm that measures a temporal EVS from the fixation offset to an articulation offset of that syllable/word.

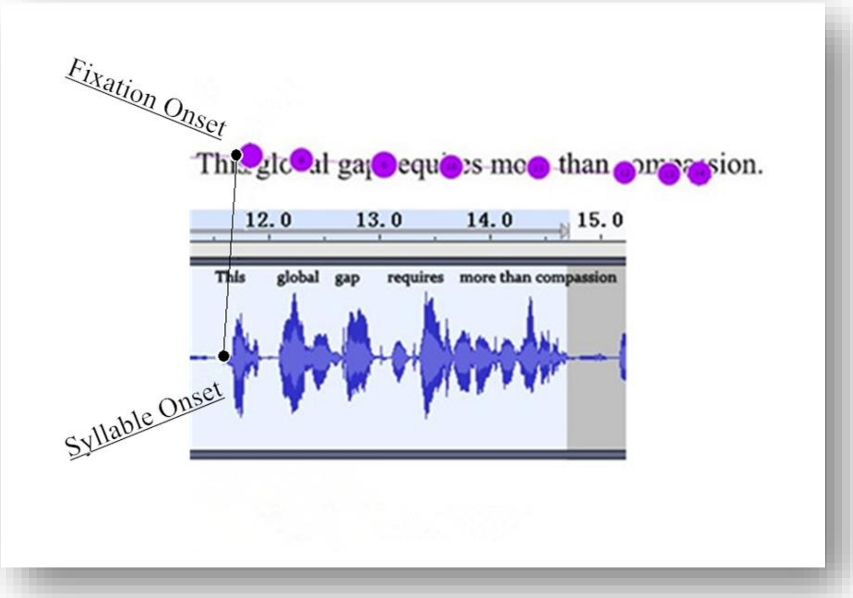


Figure 3.2a: Measuring Onset-to-onset Temporal EVS

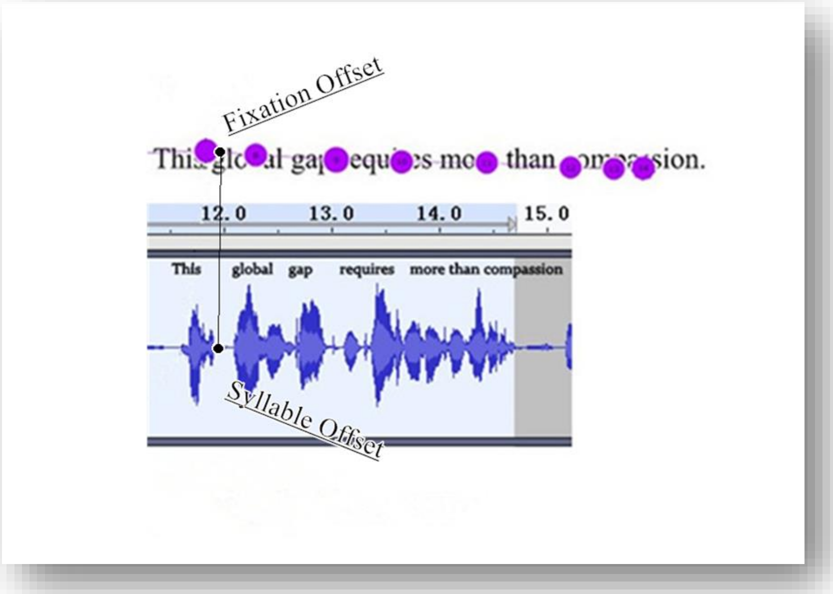


Figure 3.2b: Measuring Offset-to-offset Temporal EVS

In brief, onset-to-onset temporal EVS is measured by (the time of the articulation onset of the syllable – the time of the fixation onset). Offset-to-offset temporal EVS is measured by (the time of the articulation offset of the syllable – the time of the fixation offset). Similar to Laubrock and Kliegl's (2015: 6) finding that measuring temporal EVS from onset to onset and from onset to offset yielded completely different results, measuring temporal EVS from onset to onset and from offset to offset will also give researchers different results. To be specific, assuming the first fixation in Figure 3.2a and 3.2b lasts 200 ms and the articulation lasts 400 ms, then the onset-to-onset temporal EVS in this case is (articulation onset – fixation onset) whereas the offset-to-offset temporal EVS is [(articulation onset + 400 ms) – (fixation onset + 200 ms)]. As a result, the offset-to-offset temporal EVS is 200 ms longer than the onset-to-onset temporal EVS in this case. Therefore, these two types of calculation are not comparable. Researchers should be aware of the difference it makes if they have chosen to measure temporal EVS one method or the other. While researchers are free to choose their method, they should not mix match these two methods in a study.

A more detailed presentation of the calculation algorithm used in the present research will be presented in Chapter 5.

3.2.4 EVS in STR Studies

Temporal EVS could occur “in any task in which the same information is being treated simultaneously in two ways-by the eyes and the voice” (Levin and Addis1979: 69). Temporal EVS during STR could also “allow the mind to grasp and interpret a large unit of meaning before the voice must express it” (Buswell 1920: 41).

During RA, the readers’ attention has to be divided between the eyes and the voice (Levin and Addis 1979: 21). STR is also a task that requires the interpreters to perform reading and speaking at the same time. In the same vein, a part of the translator’s cognitive effort is detracted from comprehension to vocalisation during STR. The EVS during STR indicates that cognitive effort is devoted and divided among *Reading*, *Short-term Memory*, *Production*, and *Coordination*, the four elements of Gile’s (2009) revised *Effort Model* of STR. The dynamics between the eyes and the voice are essentially the coordination between the subprocesses. It shows that two actions are carried out simultaneously: reading new information and orally translating the previously read content. Such performance usually the result of attention division. There are at least three potential theories as to why such a division of attention is possible: because at least one of the two tasks is carried out automatically without conscious control (Hirst *et al*. 1980), because attention alternates rapidly between the two tasks (Jaffe *et al*. 1967), or because a genuine division of attention is accomplished (Downey and Anderson 1915).

Sandrelli (2003: 272), Agrifoglio (2004: 54), and Giles (2009: 180) all agree that when an interpreter is performing STR, he/she has to read ahead to identify keywords, re-structure the linguistic elements in advance, and produce smooth oral renditions of the ST. By doing so, the interpreter needs to store some information in his/her short-term memory until he/she has gathered enough information from the ST to reformulate it in the target language (Lee 2012: 695). Reading ahead in STR is similar to a reading strategy readers developed in self-paced reading, which is “a computerized method of recording a reading time for each designated segment (i.e., a word or phrase) of a sentence or series of sentences that is presented as an experimental stimulus” (Jegerski 2013: 21). Also called subject-paced reading, self-paced reading tasks present the ST in a cumulative linear format with word-by-word segmentation (ibid.). According to Ferreira and Henderson's (1990) research, many participants might reveal more than one word at a time before reading the first word aloud. It could either be an unconscious reading behaviour or a strategy applied by the reader to preview the following words so that they can deliver the sentences smoothly. The distribution and duration of temporal EVS, therefore, could be interpreted to signal the planning for production. In a way, having a temporal EVS both between and within sentences indicates that it is part of both macro-planning and micro-planning activities. While macro-planning involves the semantic and conceptual preparation between sentences, micro-planning is concerned with the cognitive process within sentence segments (Dragsted and Hansen 2008: 13).

In an eye-tracking study focused on an English-Chinese STR experiment, Huang (2011: 64) examined sentence boundaries and found that *reading ahead* occurred at 72.8% of the cases. In these cases, off-sync fixations took place at the beginning of sentences. In 26.6% of the cases there were no off-sync fixations but the reading was still ahead of the oral production. In 0.60% of the cases, however, the voice was ahead of the fixations and Huang (2011: 64) argued that it was "probably because the interpreter employed anticipation skills during interpreting and was able to predict from the context the content was coming up in the speech." Hence, in 0.60% of the cases, participants had a negative temporal EVS. These percentages indicate how common the overlap between reading and oral production is during STR, and how common it is to show a positive temporal EVS.

Research carried out by Zheng and Zhou (2018) monitored the reading ahead phenomenon around metaphorical expressions (MEs) with the help of the state-of-the-art technology. Their research focused on the temporal EVS during sight translation of metaphorical expressions (MEs) from English to Chinese by looking at the eye movements and the fixations during the pauses preceding the MEs, and during the time of translating the MEs orally. Based on previous literature, Zheng and Zhou (2018) hypothesised that temporal EVS is necessary for ensuring a smooth delivery of the TT. They specifically differentiated *Reading Ahead* around MEs into two types: Reading Ahead into

MEs (RAiM) and Reading Ahead beyond MEs (RAbM). RAiM indicates that the subjects had glanced over an ME when they were still orally translating the linguistic segments before the ME; while RAbM showed that when the subjects were processing an ME, their eye fixations had moved forward to the ensuing text. Both RAiM and RAbM were calculated by the total duration of off-sync fixations. Based on the eye-tracking data collected in this research, every individual performed the STR task with RAiM, meaning that the planning step of processing a ME took place even before the pause ahead of the ME. On the other hand, every individual also performed STR with RAbM, meaning that when processing a targeted ME, the subjects were also spending some time (or effort) on planning or monitoring the text beyond the ME. As the same type of reading ahead behaviour in STR, RAiM and RAbM were expected to have similar overall values to show that the same pace of forward reading occurs with the same individual. However, the results of the paired sample t-test ($p = 0.01 < 0.05$, $t = 2.1$) showed that the RAiM and the RAbM values had statistically significant differences. In 73.7% of cases, the value of the RAiM was higher than that of RAbM, indicating that the pace of reading ahead slowed down when a subject encountered a ME. This observation was in line with Zheng and Xiang's (2013) conclusion that MEs are processed with additional cognitive effort in the phase of reading comprehension during STR.

In conclusion, it was discovered that temporal EVS during STR indicates that the subjects' attention was divided into two different cognitive activities: perceiving new information and producing a TT with the retrieved information. Given the fact that the empirical studies on temporal EVS during RA are very few, and are even rarer in the field of STR, it is beneficial to further investigate temporal EVS using modern technologies such as eye-tracking, an accurate and preferable approach in researching temporal EVS (Zheng and Zhou 2018). The present research hopes to confirm the argument that having a temporal EVS is a very common reading behaviour during RA and STR, and that temporal EVS indicates the intensity of cognitive processing.

3.3 Revisiting EVS from a Cognitive Perspective

As discussed in the previous section, spatial EVS and temporal EVS are tightly coupled spatial and temporal aspects of the same reading behaviour (Holmqvist *et al.* 2011: 428). Temporal EVS has been studied much less intensively in the last century due to technological limitations. However, since the start of the 21st century, temporal EVS has been investigated in an increasing number of eye-tracking studies. By looking at the relationship between spatial EVS and other factors, such as reading competence and sentence segmentation, the present section intends to revisit the topic from a cognitive perspective in order to gain a better understanding of spatial and temporal EVS.

3.3.1 EVS and Reading Competence

Some studies associated the length of spatial EVS with reading competence. In Buswell's (1920) experiment on reading homographs, readers with larger spatial EVS made few mistakes. Another study of Buswell's (1921) also emphasised the need for a wide spatial EVS because a narrower spatial EVS meant readers were unaware of some forthcoming reading difficulties until it was too late. For example, the failure of immature readers to raise their intonation while reading a sentence marked by a question mark is clear evidence that reading ahead is a necessary reading skill.

Researchers have compared spatial EVS between young and mature readers in the hope to establish the relationship between one's age and the length of spatial EVS. A consistent finding is that spatial EVS increases as the reader ages (e.g., Buswell 1920; Tinker 1965; Levin and Turner 1968; Zollinger 1974). For example, Buswell (1920) reported an increasing spatial EVS in students over the course of a high-school. Various potential explanations have been proposed. According to Levin and Addis (1979), when a reader gets older, his/her short-term memory gets better, and he/she becomes better at applying his/her knowledge of grammar and semantics in reading. As an important characteristic of readers, *age* is often mentioned by researchers who want to investigate the difference between mature and inexperienced readers. However, in fact, it is the experience and skill that has the biggest impact, not age.

Buswell (1920) and Fairbanks (1937) both argued that a wide spatial EVS is a characteristic of good readers, and the average spatial EVS of good readers is wider than that of the poor readers. The foundation of such a belief is a positive correlation that has been discovered between the length of spatial EVS and reading speed, which is “an overt reflection of automaticity in decoding and similarly deft control over other component skills” (Braze 2018: 3). In other words, if a reader is more skilled and experienced, he/she might read at a faster pace and thus obtain a wider spatial EVS. Buswell (1921) summarised the relationship between spatial EVS and reading competence in an empirical study that recruited three groups of subjects with a hierarchy of reading competency: elementary school, high school, and college. The general observation showed an increase in the length of the span throughout the learning process. Also, by comparing different reader groups’ data, it was found that the poor readers do not have a variation of spatial EVS within a sentence (Buswell 1921: 223). Buswell (1921), therefore, proposed that developing reading competence is in line with expanding one’s spatial EVS. At the most primitive stage of RA, immature readers tend to have their eyes and voice focused on the same word till it has been articulated, but mature readers tend to maintain a comparatively wide spatial EVS span.

Levin and Addis (1979: 133) believed that the length of EVS only associates positively with the speed of the eyes and argued that rapid reading entails longer forward eye movements, fewer regressions, fewer and shorter fixations, thus resulting in longer EVS. They stated that experienced readers should have a longer EVS “because they have a wider attention span that allows them to grasp a larger number of elements at once” (Levin and Addis 1979: 67). However, since both spatial and temporal EVS show the dynamics between the eyes and the voice, which are interacting and coordinating during the process, they do not solely depend on how fast the eyes move. Laubrock and Kliegl (2015: 17) suggested that EVS varies in accordance with the articulatory demands. In fact, articulation is a limiting factor to the maximum reading speed (Krieber *et al*. 2017: 3), thus a limiting factor to the maximum EVS length. Although the size of spatial EVS and the length of temporal EVS could vary depending on different conditions, it has its limit because even though the eyes could, in principle, proceed faster than the voice, they need to wait for the oral articulation (Laubrock and Kliegl 2015: 17). In other words, the expansion of EVS is always limited.

Geyer (1966) contributed to the research on EVS by linking the length of it with the capacity of one's short-term memory. The argument was that one's memory capabilities limit the size of spatial EVS. Similarly, working memory capacity also determines the size of the TT produced during the simultaneous and the consecutive translation tasks (Dragsted 2004: 274). In this view, the amount of the information one can hold in the working memory during RA determines how much further his/her eyes can move ahead of his/her voice. Reading ahead allows the reader to store a certain amount of information in his short-term memory storage. Therefore, the size of the individual's spatial and temporal EVS varies (Smith 1971).

Another interesting finding is that skilled readers seemed to have a "more elastic span than poor readers" (Anderson and Dearborn 1952: 125) because they can modify their spatial EVS to fit the unit of meanings (Buswell 1920: 45). This finding contributed to moving away from the notion of having the *longest* EVS and towards having a *suitable* EVS, instead. A suitable EVS is "a consequence of higher cognitive skills" (Levin and Addis 1979: 53). The present study, therefore, does not consider *a suitable* EVS to be the *longest* EVS, no matter it is spatial EVS or temporal EVS. Specifically, developing an eye-voice span of a suitable length/duration is necessary for having an intelligent grasp of the material. Having the ability to adjust the length of EVS, both spatial and temporal, however, could be even more helpful for reading efficiently during RA and STR. Hence, whether the subject has a suitable and flexible EVS could be used to judge one's reading skill and possibly STR competence.

3.3.2 EVS and Sentence Segment

Besides being interested in the relationship between readers' reading competence and their EVS, researchers were also curious to know whether EVS has a strong connection with various factors such as task type, task demand, and position in a sentence segment.

The nature of a task has an impact on the readers' eye movements. It has also been acknowledged that spatial EVS varies from task to task and it does not have a fixed length. For instance, the average spatial EVS was said to be 3.91 words long on a normal text across different age groups (Levin and Turner 1968), whereas Gibson and Levin (1975: 363) concluded that the spatial EVS for an unstructured word-list is “short and surprisingly constant at about two words regardless of the reader’s age or ability”. The fact that spatial EVS was found to be longer when a reader reads meaning units within a paragraph rather than random words from a list (Levin and Addis 1979: 45-46) made some researchers, such as Levin and Addis (1979), realise that the fundamental research goal is to find out what influences the difference.

Judd and Buswell (1922) found that when instructed to read for details, readers would have a larger number of fixations and regressions. Regarding the relationship between reading purposes and spatial EVS, Levin and Cohn (1968) concluded that subjects had a spatial EVS of 3.97 words when they were asked to read normally; a narrower spatial EVS (3.69 words) when asked to read carefully; a wider spatial EVS (4.41 words) when asked to skim. Besides, some researchers have brought up the typography as an influencing factor on this particular phenomenon. Resnick (1970), Levin and Kaplan (1968) found out that altering the physical form of the written text, such as projecting the text up-side-down or filling up the spaces between words with symbols, significantly reduced the spatial EVS. Levin and Addis (1979: 43) thus hypothesised that the typographic abnormality “forces readers into a word-by-word strategy”. Furthermore, in a similar way to the perceptual span, spatial EVS may vary when a reader is reading different languages.

Besides task type, task demand also influences EVS. Some researchers agreed that the length of spatial EVS decreases as the reading material gets more demanding (Buswell 1920; Fairbanks 1937; Lawson 1961; Gibson and Levin 1975; Downing and Leong 1982). One example is that unfamiliar words reduce spatial EVS while familiar phrases increase spatial EVS (Quantz 1897). Buswell (1921) also pointed out that when readers encounter reading difficulties during RA, their spatial EVS reduces immediately to a primitive form. Although no consistent relationship between the size of spatial EVS and the number of regressions per line has been found in Buswell’s research (1920), Anderson and Dearborn (1952) suggested that regressions, if they occur, shorten spatial EVS.

Previously, some researchers believed that spatial EVS is likely to be determined or affected by the words' position on the line (Levin and Addis 1979: 39). Quantz (1897) claimed that the size of spatial EVS changes regularly depending on the position on the line: It is the widest (7.4 words) at the beginning of the line, and then decreases gradually (5.1 words in the middle) until it becomes the narrowest (3.8 words) at the end. Levin and Addis (1979: 40) tried to explain the pattern by arguing that "the voice 'catches up' with the eyes as they make the sweep from the end of one line to the beginning of the next". However, some other researchers approached this idea from a different angle. Instead of the position on the line, they considered the position in a sentence. The former is merely a "mechanical issue", while the latter "concerns constraints within and across sentence boundaries" (Levin and Addis 1979: 41). According to Buswell (1920), spatial EVS is as long as 12.7 letter spaces at the beginning, 12.7 in the middle, and 10.9 at the end of the line. This finding seems to be in line with Quantz's (1897) argument, but Buswell (1920) did not draw the same conclusion. Instead, the earlier finding attracted his interest in studying spatial EVS at the sentence level, and he found that subjects' average spatial EVS was 15.9, 13.4, and 10.9 letters at sentence initials, middles, and terminals respectively. Buswell (1920: 48) interpreted this finding to mean that "it is evident that the content of meaning is recognized, and that EVS is determined by thought units rather than by printed line units".

This important implication has inspired a series of studies into the grammatical structure of sentences and the length of spatial EVS. Based on Buswell's (1920) findings, researchers moved on to examine spatial EVS at the sentence level. However, the attempts to examine spatial EVS at sentence level were not always successful. For example, Fairbanks (1937) confirmed that spatial EVS was shorter at sentence terminals, but did not confirm that spatial EVS at sentence initials was larger than average. Nevertheless, O'Brien (1926), Tinker (1965), and Vernon (2014) all agreed that although spatial EVS does seem slightly narrower at the end of a line, its size has little correlation with the position on the line. Therefore, researchers started to believe that the size of spatial EVS does not have a strong association with the position on a line, but with the position in a sentence. Moreover, it was found that the more experienced readers are more able to vary their spatial EVS by sentence position whereas less-experienced readers show no such difference within a sentence (Levin and Addis 1979: 67). The implication of this finding caused discussions among researchers, but in general, they agreed that it is of considerable significance (Gibson and Levin 1975: 361).

Another significant finding from the studies of spatial EVS during RA was that the experienced readers' spatial EVS regularly narrows at phrase and clause boundaries (Levin and Addis 1979: 51). In fact, the idea that one's spatial EVS narrows at phrase boundaries was previously suggested by Schlesinger (1968) in a study exploring Hebrew reading. This idea was later supported by researchers such as Levin and Turner (1968) who confirmed this theory with a wide range of readers reading English aloud. According to Levin and Addis (1979: 97), "there is substantial evidence that grammatical phrases behave as units in various psychological tasks". Hence, phrase, clause, and sentence boundaries were suggested to be suitable units to study EVS.

In summary, a marked relationship was discovered between the length of spatial EVS and the position in a sentence: spatial EVS tends to be the widest at the beginning of a sentence, regardless of its position in the line (Buswell 1921: 221). Readers tend to read further ahead at sentence initials to grasp as much information as they can, and slow down at sentence terminals to let the voice output catch up with the reading input. The fact that the size of one's spatial EVS varies within a sentence and follows a pattern shows that spatial EVS is "determined by thought units" (Buswell 1920: 50). If this was not the case, the partial EVS would be mostly consistent or completely random. Since spatial EVS and temporal EVS are "tightly coupled aspects of relative motion" between the same set of visual and vocal events (Holmqvist *et al.* 2011: 428), and because of the lack of effort made to distinguish these two measures, the present research proposes to investigate temporal EVS in sentence segments to discover whether temporal EVS is also the longest at sentence initials and the shortest at sentence terminals. If so, this shows that the temporal aspect of EVS, like its spatial counterpart, also indicates that the intensity of cognitive processing is not evenly distributed across a sentence.

3.3.3 EVS as an Indicator in Empirical Studies

Since “eye-tracking data are very versatile; rich in information in both the spatial and the temporal domains” (Holmqvist *et al*. 2011: 454), researchers can choose the indicators that are the most appropriate for their research. Recently, Carl and Schaeffer (2016) pointed out that researchers should not settle with using only the traditional eye movement measures. There are other indicators of cognitive load in translation besides fixation-based measures, pupil-based measures, and saccade-based measures (Hvelplund 2014: 215). Indeed, researchers have the freedom to employ different measures and indicators in their studies. Moreover, researchers should also be encouraged to develop new indicators such as latency measures, which is among less conventional but no less informative indicators.

Dragsted and Hansen (2008) introduced eye-key span in translation studies, which is a latency measure referring to the time span between casting the first fixation on a specific word in the ST and typing out its TT equivalence. They carried out the study with the eye-key span in written translation, aiming to test whether pauses are indeed true indicators of the boundaries of translation units (Dragsted and Hansen 2008: 12). Their analysis suggested that words and phrases may have attracted the translators’ attention long before they were translated and even before the preceding pauses took place. Meanwhile, the segments that are more difficult showed relatively longer temporal eye-key spans. In short, the length of temporal eye-key span should relate to the coordination of the reading comprehension and the typing production during written translation. Therefore, it could be hypothesised that temporal EVS during RA and STR also signals the intensity of the integration of subjects’ reading input and oral output. The question here is whether a longer or a shorter temporal EVS represents the better coordination of the cognitive events during reading and speaking.

As discussed before, temporal EVS is a latency measure that fits the definition of latency measure in eye-tracking research. Temporal EVS is certainly not an error like a system latency caused by the eye-tracking programme’s lack of accuracy and precision (Holmqvist *et al*. 2011: 428). Inspired by Dragsted and Hansen’s (2008) research, the present research aims to explore temporal EVS as another possible indicator of cognitive load and cognitive management in RA and STR. Temporal EVS shows the existence and the extent of parallel processing. As an analysis tool, it was relatively well developed and mapped out within the paradigm of RA, but this does not prohibit us from using this measure to tackle research questions in the area of STR.

In comparison, spatial EVS has been investigated more intensively in reading studies. The spatial length of EVS has a positive correlation with reading speed, and it has a positive correlation with reading competence. Although its size is limited by one's working memory capacity, spatial EVS can be developed gradually. Nevertheless, what was missing from literature in this field is whether temporal EVS works in the same way. As the temporal counterpart of spatial EVS, temporal EVS seems to be overlooked sometimes. Because of the lack of distinction between spatial and temporal EVS in literature, researchers tend to make conclusions about temporal EVS based on their investigation on spatial EVS.

In more recent years, however, some researchers have focused on the temporal domain of EVS. For instance, Timarová *et al*. (2015) found shorter temporal EVS for more experienced interpreters. In another study of the eye-key span as a measure for translation complexity, Carl and Schaeffer (2016) supported Dragsted's (2010) finding that professional translators have a shorter temporal eye-key span than student translators, which was measured by deducting the fixation onset on a word in the ST from the typing onset of the equivalent word in TL. These temporal measurements yielded findings that were different to findings yielded by their spatial counterparts. However, this is not surprising. Temporal measurements, such as second and millisecond, are fixed measures of time and can be used across different domains for brain activity, whereas spatial measures such as syllable and word vary according to the spatial context. This fundamental difference between spatial and temporal EVS means that temporal EVS and spatial EVS should be used and analysed separately. The lack of focus on temporal EVS in literature encouraged the present study to investigate only temporal EVS to generate more accurate results using eye-tracking.

The present study aims to achieve a direct measure of temporal EVS during RA and STR with the help of the eye-tracking software and its detection algorithm and, therefore, does not encourage comparing either spatial or temporal EVS calculated through different methods. The lights-out technique (Levin and Kaplan 1970) gave the researchers "an indirect measure of the latency between reading and speaking" (Holmqvist *et al*. 2011: 444), and thereby covering the text or turning off the light has been occasionally criticised for being a simple guessing game (Gibson and Levin 1975: 368). The up-to-date eye-tracking methodology, however, brought about a more scientific and accurate measure of this indicator. Holmqvist *et al*. (2011: 444) discovered a huge difference between the EVS result from Sloboda's (1985) study and Goolsby's (1994) study. The former applied the lights-out technique and the latter employed an eye-tracking device. Holmqvist *et al*. (2011), therefore, highlighted the problem of comparing the EVS results obtained from different studies that used different measuring techniques.

By making a thorough comparison between temporal EVS and other major eye movement indicators, this research will test the potential of temporal EVS as an indicator of cognitive effort devoted to reading and translation. The present study suggests that STR research could benefit from exploring this measure, which has been rarely used in translation process studies. Since ST perception during translation differs from that during a non-translation-specific situation (Göpferich 2009: 14), RA and STR are undoubtedly different tasks. The present research, however, hopes that temporal EVS could become an important parameter for investigating both processes. Furthermore, while the vast majority of the existing research focused on studying spatial and temporal EVS that takes place during RA in English, it is acknowledged that "the grammatical and semantic constraints in different languages provide natural variations for the study of the EVS" (Levin and Addis 1979: 43). Cross-language comparisons of EVS have been carried out with English, Hebrew, and Japanese, but the findings are limited. However, comparing spatial EVS in reading English and languages such as Chinese and Japanese is interesting as the writing units are entirely different in their orthographic system. Specifically, there are no letters in the Chinese writing system and word-length in Chinese always remains the same. Consequently, the present research hopes that temporal EVS can yield more valuable results when the reading materials are in different languages.

Chapter 4. Research Methodology

Cognitive linguistics investigates “the relationship between human language, the mind and socio-physical experience” (Evans, Bergen, and Zinken 2007: 2). Founded by cross-disciplinary empirical methodologies, the study of cognitive linguistics initially emerged in the 1970s (Fillmore 1975; Lakoff and Thompson 1975; Rosch 1975) and has been significantly influenced by theories and findings from other branches of cognitive sciences, especially cognitive psychology (Evans, Bergen, and Zinken 2007: 2). Two critical commitments of cognitive linguistics are the *Generalisation Commitment*, which aims to characterise the general principles that apply to all aspects of human languages, and the *Cognitive Commitment*, a set of general principles for languages incorporating knowledge of the mind and the brain from other disciplines (Lakoff 1990).

Innovative research tools have attracted an increasing amount of attention from cognitive linguistics researchers (Saldanha and O’Brien 2014: 109), who have stated the need for a valid and reliable data elicitation method. For example, translation recognition tasks have been applied to measure participants’ Reaction Time (RT) when they match words and their translation equivalents in two languages. The underlying assumption is that “longer RTs and greater error rates indicate a processing difficulty” (Sunderman 2013: 188). It was argued that “researchers could get a glimpse of what participants were activating in their minds and intending to produce” (ibid.). Translation process also attracted interest from those researchers in the field of cognitive linguistics. As a task that involves comprehension, processing, and production of languages, translation is no longer just used as a type of experimental task. Translation has become a research subject.

Therefore, since the early 1980s (Hvelplund 2011: 10), a wide range of research methods have been applied to investigate cognitive processing during translation in a series of empirical studies. In the 21st century, various research methods have been used to study the translation process, such as Think-Aloud Protocols (TAPs) (e.g., Zheng 2012a; Zheng 2012b), audio-recording (e.g., Dragsted, Mees, and Hansen 2011; Lee 2012), quality rating (e.g., Lambert 2004; Dragsted and Hansen 2009; Pedersen and Dam 2014), speech recognition software (e.g., Gorszczyńska 2010; Dragsted, Mees, and

Hansen 2011), retrospective interview (e.g., Zheng and Xiang 2013; Zheng and Zhou 2018), corpus-based methodologies (e.g., Shlesinger and Ordan 2012), eye-tracking (e.g., Pavlović and Jensen 2009; Huang 2011; Alves, Gonçalves, and Szpak 2012; Chmiel and Mazur 2013), keyboard logging tools, e.g., *Translog*, (e.g., Jakobsen 2003, Dragsted 2005, Dragsted, Mees and Hansen 2011), and methods more frequently applied by neuroscientists, including electroencephalograph (EEG), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET) (Göpferich and Jääskeläinen 2009: 170).

When studying the translation process, different research methods yield different types of data, which provide information about different aspects of the translation process. When considering different research methods, validity and reliability are the main concerns of the researcher (Göpferich and Jääskeläinen 2009: 179). While the abovementioned empirical studies aimed to collect accurate and truthful data, they also attempted to minimise any influence from their experimental settings. In many cases, different methods have been combined to give researchers robust and comprehensive data to investigate the translation process. For example, key-logging and eye-tracking are often combined in studies that focus on written translation (e.g., Dragsted and Hansen 2009; Jakobsen 2011; Carl and Dragsted 2012; Sjørup 2013; Hvelplund 2014). Although all the research methods have their strengths and weaknesses, researchers need to select the most appropriate research methods depending on their specific research questions and aims. The choice of using a particular research method is "a compromise between a number of factors, including validity, reliability, and the availability of subjects and resources" (Göpferich and Jääskeläinen 2009: 171).

Eye tracking-based empirical studies focus their analysis primarily on eye-tracking data, which records the movement of the eyes during reading. Eye movements are under oculomotor control (Downing and Leong 1982: 144). Two types of eye movements are of great importance during reading: the changes in the pupil size and the eye-balls' movements (Downing and Leong 1982: 141). *Pupil dilation* is the result of the first type of eye movements, whereas *fixations* and *saccades* come from the second type. In the earlier stage of eye-movement studies, researchers focused on the oculomotor aspects of processing, but in the 1970s, the study of eye movements and reading began to be more concerned with the information processing of text materials (e.g., Gaarder 1970). The fundamental hypothesis, nevertheless, is that "there is a correlation between eye movements and pupil dilation on the one hand and the perceptual and cognitive processes going on during these eye movements on the other" (Göpferich and Jääskeläinen 2009: 173).

4.1 Eye-tracking and Audio-recording

The present research will combine eye-tracking data with audio-recording data, with the aim of providing solid and comprehensive data to investigate the STR process. This research methodology gives the researcher a valid means to observe the way in which the input and output proceed and coordinate with one another.

4.1.1 Eye-tracking

4.1.1.1 The Most Popular Types of Eye-trackers

Eye trackers are "advanced psychological measuring systems" (Holmqvist *et al*. 2011: 10). According to Holmqvist *et al*. (2011), the earliest eye trackers were built in the late 1800s. Later, Dodge and Cline (1901) introduced the principle of photographing the reflection of an external light source from the fovea and video-based pupil-corneal reflection tracking became the dominant method in the 1990s (Holmqvist *et al*. 2011: 25). A number of eye-tracking techniques have been developed by individual researchers since 1950: lens systems, electromagnetic coil systems, electrooculography systems, and dual Purkinje systems. In the mid-1970s, eye-tracking equipment became more accessible from engineering companies and, therefore, researchers no longer had to build their own hardware and resolve any technical issues (Holmqvist *et al*. 2011: 10).

In general, there are three types of eye-tracking devices: head-mounted, head-supported, and remote eye trackers. Different researchers must select one of the three types of eye-tracking devices based on their research purposes and conditions. The head-mounted system is quite "invasive" because the subjects have to wear the equipment on their head (Hvelplund 2014: 263). However, the advantage of this type of system is that "the recording area of the device is not restricted to the screen area of a computer monitor" (Hvelplund 2014: 206). Secondly, the head-supported eye-tracking device is usually chosen to achieve higher accuracy, which is an important consideration in deciding which eye tracker is the best for a given experiment. Its ecological validity, however, is questioned (Saldanha and O'Brien 2014: 139). Finally, there is the remote eye-tracking device, also referred to as a desktop eye tracker due to its position. The cameras are usually "integrated into a separate box which is placed in front of or attached to, a computer monitor" (Hvelplund 2014: 205). The remote eye tracker has been favoured in many translation studies that aimed to imitate an authentic research setting as it is the least invasive type of the three kinds of eye trackers available (O'Brien 2009: 263). O'Brien (2009) argued that this type of desktop eye tracker increases the ecological validity of the research process. However, some researchers acknowledged the main disadvantage of using a remote eye tracker by pointing out that "the level of eye tracking quality in terms of spatial accuracy is lower (up to 1 degree of inaccuracy) than that of a head supported tracker" (Pavlović and Jensen 2009: 97).

Given that all three types of eye trackers have their advantages, different researchers should select which eye tracking device to use depending on their experimental focuses and task design if they can. In the present research, the merit of the head-mounted eye-tracking device is not advantageous because the reading tasks (each lasting 120 seconds) and the STR tasks (each lasting 200 seconds) are much shorter than written translation tasks. Moreover, subjects are much less likely to look away from the screen because there is no need to look at the keyboard during RA and STR. The use of a head-supported eye-tracking device would not work well in this study as the subject's chin would be supported. The use of a chinrest would be problematic as jaw movements during RA and STR tasks have a considerable impact on the position of the eyes when the subject speaks. Thus, the use of a head-supported eye tracker would have a negative impact on the quality of the eye-tracking data as jaw moments would be limited by the chin being supported. Due to this issue, Krieber *et al*. (2017) removed the chinrest in their RA experiments to facilitate oral production. However, the participant having freedom to move meant that the data quality and accuracy was compromised.

As Saldanha and O’Brien (2014: 142) pointed out, the remote eye trackers are not “consistently 100% accurate in the capture of gaze data”. However, researchers should be aware that 100% accuracy cannot be guaranteed when using any research method. Some participants will have lower or higher accuracy than others and a single individual’s recording may have lower or higher accuracy from time to time. Based on this comparison, the present research considers a remote eye-tracking device attached to a regular monitor as the best option to preserve the ecological validity and to minimise the *white-coat effect* (Saldanha and O’Brien 2014: 118). Data quality and accuracy will be preserved by carrying out a thorough data quality assessment, which is elaborated on in chapter 5.

4.1.1.2 Eye-tracking in Applied Linguistics Research

Eye tracking is “an experimental method that consists of monitoring and recording the eye movements that a person makes while performing a task that involves complex visual cognitive processing” (Keating 2013: 69). It is “the process of recording the point of gaze of a person and the movement of the eyes from one point to another” (Saldanha and O’Brien 2014: 136).

When researchers discuss eye-tracking methodologies now, they think of the use of eye trackers. Before the age of eye trackers, however, there were a variety of other techniques used to record eye movements during reading and other visual tasks (e.g. Ahrens 1891; Sperling 1960; Bouma and de Voogd 1974; Gibson and Levin 1975; Rayner and Pollatsek 1989). Traditional eye movement recording methodologies have used electro-oculography, the corneal reflection method, the contact lens method, and tracked the limbus, pupil, and eyelid (Downing and Leong 1982: 145). After photographic techniques were produced, researchers were able to record the reflection of a beam of light directed at the reader’s eyes on photographic films. Although some people were sceptical about this method because it immobilised the subjects’ heads, this method was “the major technological advance” in studying reading process (Gibson and Levin 1975: 354). The number of researchers who have used professional eye-tracking devices to record visual processes in different research fields has grown enormously since 1990 (Holmqvist *et al*. 2011: 10).

An eye-tracking device is capable of eliciting eye-tracking data, such as a participant’s fixations duration, fixation counts, saccadic, and papillary movements. Besides the electrical activity of the brain, eye movement is another, if not the only, valuable and measurable event “that provide indicators of ongoing processing of information during reading with minimal interference of the process” and is capable of reflecting “cognitive processes that operate at different levels of abstraction and that occur simultaneously” (Wolverton and Zola 1983: 41). Saldanha and O’Brien (2014) argued that eye trackers are perhaps best used for eliciting data concerning attentional and cognitive effort in translation process studies. The primary advantage of the eye-tracking technique, according to Frazier (1983: 221), is that it permits the researchers “to obtain evidence about what is happening during the comprehension of a sentence without significantly altering the normal characteristics of either the task or the presentation of materials”. Three decades ago, Ehrlich (1983a: 254) correctly predicted that the eye movement recording technique had “the potential of being extremely useful” because it can “provide much fine-grain data while minimizing any artificiality” (ibid.). The recent theoretical and technological advances in the use of eye-tracking methodologies have considerably enhanced the study of the cognitive processes during translation. This technique can provide a moment-by-moment record of where the eyes are looking, for how long, and the way they move. Much progress has been made in mapping out many of the cognitive operations in translation-based research on the quantitative data yielded by eye tracking.

Modern-day eye movement research can be traced to the mid-1970s (Rayner 1998: 372). Since then, as a powerful quantitative and qualitative research method, it has provided enormously valuable information in reading studies (e.g., Rayner and Pollatsek 1989; Rayner 1998; Jakobsen and Jensen 2008; Pellicer-Sánchez 2016), and it has been adopted by translation-process researchers. In 2006, O’Brien (2006) still viewed eye tracking as a methodology rarely applied in translation studies, but the field has changed a lot since then. Recently, Walker (2018) pointed out that translation process research has witnessed an increasing use of eye-tracking methodologies in the past decade. Specifically, eye-tracking methodologies have been used in a broad range of studies on written translation (e.g., O'Brien 2009; Pavlović and Jensen 2009; Jensen, Sjørup, and Balling 2009; Jakobsen 2011; Hvelplund 2015) and STR (e.g., Dragsted and Hansen 2009; Shreve *et al*. 2010; Korpal 2012; Chmiel and Mazur 2013; Zheng and Zhou 2018). It has been applied to address many other cognitive-related translation topics such as translation memory (O'Brien 2006), metaphor translation (Sjørup 2008), reading for translation as a special kind of reading (Jakobsen and Jensen 2008; Hvelplund 2015), coordination between comprehension and production during translation (Dragsted and Hansen 2008), translation directionality (Pavlović and Jensen 2009; Chang 2009; Wang 2017), and translators' competence (Ehrensberger-Dow and Massey 2013).

4.1.1.3 Eye-mind and Immediacy Assumptions

Three pairs of muscles control human eye movements, and they are responsible for the "horizontal (yaw), vertical (pitch), and torsional (roll) eye movements" respectively and thus "control the three-dimensional orientation of the eye inside the head" (Holmqvist *et al*. 2011: 21). Since "large parts of the brain are engaged in controlling these muscles" to "direct the gaze to relevant locations in space" (ibid.), one's eye-movements and mind are connected and cooperate in a sophisticated way. Psycholinguists have been interested in understanding the relationship between the eyes and the brain. In order to study this relationship, researchers have tried to analyse the structure of the eye and the brain. For example, Penfield and Roberts' (1959) found a link between a certain point on the retina and its corresponding part in the cortex. Their finding has been confirmed by researchers such as Hubel and Wiesel (1962), who experimented on the visual cortex in animals.

The link between cognitive effort and eye movements is one of the most interesting aspects of the eye-tracking methodology used in translation studies (O'Brien 2006: 186). Many researchers interested in translation-process research, such as Jakobsen (2011) and Hvelplund (2014), agreed that the use of the eye-tracking method in empirical research is based on Just and Carpenter's (1980) *eye-mind assumption* and *immediacy assumption*. These two assumptions state that the eyes remain fixated on what was processed at the time during reading, and that readers start their interpretations immediately after they encounter the written words. Their assumptions have been widely used as an "operational basis for assuming a link between visual focus and cognitive focus" (Hvelplund 2014: 209). In Jakobsen's (2011: 38) words, "the fundamental assumption here is that there is a correlation between behavioural 'outside' data and cognitive 'inside' processing".

Many researchers have pointed out a strong connection between the location of a fixation and the content being processed (e.g., Rayner and McConkie 1976; Posner 1980; Anderson 2000). Even though Just and Carpenter's (1980: 331) eye-mind and immediacy assumptions are "used as an operational basis for assuming a link between visual focus and cognitive focus" (Hvelplund 2014: 209), these hypotheses were questioned (e.g., Hyönä, Lorch, and Rinck 2003). Some have claimed that there was a processing lag behind the fixations (e.g., Bouma and de Voogd 1974; Kolers 1976; Morton 1964). For example, Hvelplund (2014: 209) pointed out that "the reliability of eye-tracking data as indication of cognitive processing has not yet received much critical attention in the context of translation research". Nevertheless, it is worth noting that Rayner (1983: 107) concluded that "much of the visual information necessary for reading can be acquired during the first 50 ms of a fixation". To some extent, this argument supported Just and Carpenter's (1980) immediacy assumption.

In his attempt to study written translation process, Jakobsen (2011: 38) pointed out that the eye-tracking method "radically improves our chances of reconstructing both the comprehension processes that precede production and the way in which comprehension and production processes combine". For STR, where there is no typing involved, eye-tracking data seems even more valuable. Therefore, although the eye-mind and immediacy assumptions are not perfect, they still "offer a reasonable basis for assuming some sort of relationship between eye movements and translation processing" (Hvelplund 2014: 211). Furthermore, most researchers agree that "what the eye is looking at is (in general) something the mind is attending to" (Jakobsen 2011: 47). Nowadays, the eye-mind and immediacy assumptions are widely acknowledged as "reasonable assumptions that are not only necessary in order to be able to interpret eye movements as correlates of cognitive processing in translation but that have been successfully validated in neighbouring research disciplines" (Hvelplund 2014: 211).

4.1.1.4 Advantages and Problems of Eye-tracking Methodologies

In 2006, O'Brien (2006: 186) emphasised the advantages of using an eye-tracking methodology as it "offers a particularly interesting addition to some of the methodologies already used in translation process studies". In recent years, eye-tracking methodologies have been frequently studied, examined, and applied to investigate the cognitive processes involved in different activities. One of the inherent advantages of using an eye-tracking methodology is the wealth of data it yields. It has proved to be well-established and is increasingly popular in different fields (Hvelplund 2014: 203). Indeed, studies within the field of psychophysics, cognitive neuroscience, and computer science have all employed this research method as a complementary methodology for other research tools (Duchowski 2003). Furthermore, thanks to constant technical developments and methodological progress, adopting an eye-tracking methodology has gradually developed into an interdisciplinary activity (O'Brien 2006: 185).

In general, eye-tracking tools can show the researcher the subjects' input process very precisely and reveal what the subjects might not be aware of, such as some reading strategies and problem-solving mechanisms. It is not, however, free from potential drawbacks and problems. Some researchers have discussed various methodological issues and challenges involved in the use of eye tracking in translation research (e.g., O'Brien 2009; Alves *et al.* 2009; Hvelplund 2014). The biggest challenge, as Hvelplund (2014) pointed out, is that the subjects' thoughts could *drift* due to fatigue or distraction. More specifically, *drift* refers to the situation when "the recorded eye position and the true eye position become gradually asynchronous as a data-collection session progresses" (ibid.: 211). However, drift is less likely to be a significant problem in the present study's RA and STR experiment because it is less likely to occur in short translation sessions (ibid.: 209). It has also been hypothesised that "part of the cognitive processing may not take place during the first fixation on a word" (Sjørup 2008: 59). Nevertheless, not many eye tracking-based studies have chosen to exclude the first fixation, which is also called the "first pass" (ibid.).

Another potential challenge for employing the eye-tracking method in an artificial experimental situation is how much ecological validity and reliability can be achieved (Göpferich and Jääskeläinen 2009: 171). There are questions about the ecological validity and reliability of this experiment as it combines eye-tracking data with audio-recording data. First of all, the question of the ecological validity of the RA and STR experiments is whether the research design has changed the ordinary state of the two tasks. According to Göpferich and Jääskeläinen (2009: 182), "ecological validity is a concern of all experimental studies, in which the normal situation is always somehow manipulated". More specifically, is the manipulated experimental situation changing the characteristics and phenomenon that researchers set out to investigate in the two tasks? In order to not alter the research object, which is subjected to experimental control, the research design has focused on maintaining a non-intrusive environment. In comparison, the remote eye-tracking equipment used in the present study is less intrusive than other types of eye trackers, which increases the ecological validity of its use. Secondly, a question frequently raised about the reliability of RA and STR experiments is whether researchers can trust the data they obtain through the research methods. After all, being able to reflect on the objects of the research both accurately and truthfully is extremely important. Since eye-tracking studies are based primarily on quantitative data, specifying and demonstrating how the data was collected and analysed is a crucial step in ensuring the reliability of the results. Researchers need to acknowledge and explain the constraints and limitations of their research designs to ensure the results are reliable (Göpferich and Jääskeläinen 2009: 182). Furthermore, the field of process-oriented translation studies requires additional methodological research to determine the validity and reliability of eye-tracking methodologies.

In addition, it is impossible to achieve perfect data quality. Holmqvist *et al*. (2011: 118) summarised the possible optic conditions that might endanger eye-tracking data quality: issues caused by droopy eyelids; confusion caused by mascara, glasses, ambient infrared light, retinal reflection, specks, and dirt; distortion caused by bifocal glasses; data loss caused by head movements. Among them, glasses and contact lenses are commonly the cause of most problems. Glasses can, for instance, reduce the contrast between the pupil and the iris and thus reduce the accuracy and precision of the data, or interfere with the pupil and the corneal reflection by shadowing the eyes. Soft contact lenses, on the other hand, can also affect the eye tracker's ability to track the pupil and the corneal reflection due to potential optic artefacts resulting from small air bubbles gathered underneath the lenses (Holmqvist *et al*. 2011: 122-124). There is no easy solution to these potential problems because if a subject has droopy eyelids or thick glasses, the researcher cannot change it immediately. Nevertheless, the present research suggests that there are several ways to avoid the trouble: a) eliminating subjects who cannot see the stimulus on the screen unless they are wearing very thick glasses; b) encouraging near-sighted subjects to wear contact lenses during the experiments; c) developing a series of data quality measures to identify potentially deviant subject/data.

In summary, despite the very few concerns stated above, translation-process research is confident that the "observable, quantifiable gaze data which we can collect with an eye tracker are indicative of underlying, not directly observable cognitive processes that take place during a certain task" (Jessen, Sjørup, and Balling 2009: 322).

4.1.2 Audio-recording

Audio-recording, as the largest complimentary data source to eye-tracking data (Holmqvist *et al*. 2011: 96), is commonly used to record both the process and the product of oral interpreting and STR. Researchers such as Dragsted, Mees, and Hansen (2011), and Lee (2012) recorded subjects' STR process, then analysed their performance and assessed the products. Audio data is not only the most investigated and the most well-used auxiliary data to disambiguate eye-tracking data (Holmqvist *et al*. 2011: 66) but is also an essential research method in studies investigating EVS in RA (e.g., Buswell 1921).

Researchers might choose to only examine the transcription from an interpreting or STR task, but listening to the audio data gives the researcher a better chance of understanding the process thoroughly. However, relying solely on audio data to study RA or STR processes might raise some concerns. For instance, researchers who relied solely on audio-recording subscribed to the idea that a pause is a behavioural reflection of cognitive process (Schilperoord 1996: 9) and it has a strong relationship between the amount of cognitive effort used in planning oral production (Butterworth 1980: 159). When combined with the eye-tracking method, however, it was shown that "the automatic pause-based criterion for segmentation may not be the most appropriate" (Dragsted and Hansen 2008: 25). Furthermore, Zheng and Xiang (2013, 2014) applied the audio-recording without eye-tracking method to analysis processing metaphorical expressions (MEs) during English-Chinese STR. However, they addressed the potential deviation in their articles: the planning step could have started prior to the pause, which means that it is possible that the translator may have already started processing the ME even before he/she has finished translating the segment situated before the ME in the ST. Thereafter, Zheng and Zhou (2018) combined eye tracking and audio-recording to revisit the calculation method for processing time during the sight translating of metaphors. After comparing eye-tracking data and audio data, Zheng and Zhou (2018) concluded that the mean percentage of methodological deviation is 9.58%. Since the values retrieved by the two different methods do not show statistically significant differences, the use of audio data was proven to be broadly valid. Nevertheless, these studies showed that relying only on either eye tracking or audio-recording to investigate RA and STR processes is not preferable. These examples indicated that audio-recording should be combined with an eye-tracking methodology.

In addition, there are two other common audio-recording tools: screen recordings and video recordings. The former documents everything that occurs on the computer screen and the latter records not only the audio tracks but also the visual images. However, screen recording is not employed in the present research because, unlike written translation, there is no translator action on the screen during STR. Although Göpferich and Jääskeläinen (2009) advocated the application of video recording, which might give the researchers access to the subjects' facial expressions and body language, it also might make the subjects uncomfortable and too self-conscious about their appearance or nervous about the experiment setting. In order not to sacrifice the ecological validity of the data, the present research used an audio-recording device to investigate RA and STR processes.

In addition, eye tracking has been combined with a variety of verbal report methods, including concurrent protocols such as Think Aloud Protocols (TAPs) and retrospective verbalisations (Göpferich and Jääskeläinen 2009: 171). However, neither concurrent verbal protocol nor retrospective verbalisation works perfectly in TPR. On one hand, concurrent protocols interfere with the translation process. On the other hand, the reliability of retrospective verbalisation is jeopardised greatly by memory failure (Göpferich and Jääskeläinen 2009: 181). Researchers, such as Russo *et al.* (1989) and Ericsson and Simon (1980), also noted the issue of using verbal protocols due to there being a risk of fabrication. It should be noted that the participants' verbal report might not be entirely consistent with their actual behaviour (Ericsson 2010: 247). Furthermore, verbalisation was not used during present research, as it was not very useful or accurate when it was applied to investigate EVS in the pilot study.

4.2 Eye Movements and Eye-tracking Indicators

Pupil dilation has been acknowledged by many researchers as an indicator of an increased amount of cognitive effort being devoted to a visual task (e.g., Hess and Polt 1964; Marshall *et al.* 2003; O'Brien 2008). *Fixations* and *saccades*, on the other hand, are oculomotor events that are countable entities in eye-movement data (Holmqvist *et al.* 2011: 2). Rayner and Pollatsek (1989) proposed that although the underlying mechanism of the control of eye movements is the same, there are different patterns. There are a large number of fixation-based indicators. Total fixation duration is a very common indicator used to measure the amount of the cognitive effort devoted to completing a translation task (e.g., Pavlović and Jensen 2009; Hvelplund 2014). Although eye-tracking studies applying saccade-based indicators are rare, there are still a series of saccade-based indicators. Researchers who specialise in eye-tracking research have been trying to establish and apply different eye movement-based indicators. It is very important to have a good understanding of these eye-movement indicators because they are the foundation of eye-tracking methodology and most eye-tracking based studies.

4.2.1 Pupil Diameter

The pupil is the aperture of the eye. When the light comes in through a pupil, images are turned upside down in the lens and then projected onto the retina situated at the back of the eyeball. Pupil dilation is subject to the modification by a light reflex and by muscles as it increases the amount of light entering the eye (Downing and Leong 1982: 136). Since the 1960s, many cognitive psychology studies have studied the behaviour of the pupil (Caffrey 2008: 129). The pupil has been recognised by researchers for many decades as an indicator of increased cognitive effort being devoted to visual processing (e.g., Hess and Polt 1964; Marshall *et al*. 2003; O'Brien 2008).

Researchers in the field of translation and interpreting studies confirmed that the pupillometry is a valuable assessment of the cognitive load during the translation process (e.g., Hyönä, Tommola and Alaja 1995; O'Brien 2006; Dragsted and Hansen 2007; Caffrey 2008; Chang 2009; Pavlović and Jensen 2009). A firm correlation has been drawn between cognitive effort and pupillary responses by various studies: pupil dilation increases with cognitive load (e.g., Hess and Polt 1964; Nakayama, Takahashi and Shimizu 2002). In a recent study, Iqbal, Adamcyzk, Zheng, and Bailey (2005: 312) suggested that "pupil size is an effective and reliable measure of mental workload, where increases in pupil size correlate with increases in mental workload". In other words, if the pupils are more dilated, it indicates that more cognitive effort has been devoted to the task; if the pupils are less dilated or constricted, it indicates that less cognitive effort has been dedicated to the task.

The percentage of pupil dilation, also called *PCPD value* (O'Brien 2006: 191), can be calculated by contrasting a subject's average pupil size in a baseline task with his/her average pupil size in another task. For example, O'Brien (2006) used a reading task as the baseline task when investigating measures of pupil dilation. She argued that it offers the researchers a relatively sound way of comparing pupil dilations during different reading tasks. Subsequently, Saldanha and O'Brien (2014: 144) argued that pupil dilation can be a more accurate indicator when it is recorded first in a relevant baseline task and then in a comparative task. They believed that comparing a subject's pupil dilation in other ways might be less reliable. Moreover, using the percentage of Pupil Dilation (PCPD) in cross-group comparisons should consider the ethnic origin of the subjects because pupil size might be an additional variable. Another important reason not to conduct cross-group comparisons with pupil size and PCPD value is that eye colour may affect eye trackers when differentiating the pupil and the iris (O'Brien 2009: 257).

PCPD value could be calculated by noting down the constant changes in the subject's pupil size. However, this method is somewhat problematic because it has been reported that there is a short delay in pupil dilation during reading. When pupil dilation is in response to a stimulus, it usually occurs with some delay, which is called *pupillary latency* (Hvelplund 2015: 214). Some researchers have strived to measure this pupillary latency, and the results were generally below half a second: it was reported to be 100-200 ms by Saldanha and O'Brien (2014: 144) and 300-500 ms by Hyönä *et al*. (1995: 605). Thus, should PCPD value be employed to investigate the RA and STR processes at the micro level, for example, at the lexical and or the morphologic level? Researchers should take the possibility of the pupillary latency into account. Nevertheless, PCPD value is still a valuable and accepted measurement. For instance, in Pavlović and Jensen's (2009) research, the pupillometric indicator was consistently successful in testing all five hypotheses and was mostly found to be statistically significant, while the other indicators failed to exhibit statistical significance.

4.2.2 Fixation-based Eye-tracking Indicators

Referred to as *pauses* by Gibson and Levin (1975), fixations are "eye movements which stabilize the retina over a stationary object of interest" (Duchowski 2003: 43). The purposes of fixations are the following: to transmit visual stimuli when the eyes remain relatively still, to avoid interfering stimuli during the period, and to allow comprehension time (McConkie 1983: 76).

According to Downing and Leong (1982: 142), the two neurological mechanisms involved in the control of fixations are the *voluntary fixation mechanism* and the *involuntary fixation mechanism*, which are controlled by different areas of the cortex. The first mechanism is responsible for moving the eyes to search for an object to fix one's vision on and the second mechanism is responsible for locking the vision on to an object. The voluntary searching movements of the eyes are relatively simple and they are conjugated by rapid saccadic movements. It might be assumed that it is easy to fix one's vision on a single object, but in fact, human eyes are "imperceptible in continuous movement" (Downing and Leong 1982: 142). There are three distinct types of intra-fixation micro eye movements: *tremor*, a result of imprecise muscle control; *drift*, a slow movement that takes the eye away from the centre of a fixation; and *micro-saccade*, a quick movement that brings the eye back to the centre (Holmqvist *et al*. 2011: 22-23). All these eye movements during reading "serve the purpose of locating appropriate data to increase available information" (Downing and Leong 1982: 143).

The duration of fixations varies depending on the task (Rayner 1998: 373). A fixation could last "anywhere from some tens of milliseconds up to several seconds" (Holmqvist *et al*. 2011: 22). The typical duration of fixations could be between 100 ms and 1000 ms (Sharmin *et al*. 2008: 46). Researchers have measured the duration of fixations in various experiments. For example, Gibson and Levin (1975: 356) claimed that a reader's fixation duration is between 240 ms and 250 ms when reading easy texts. Downing and Leong (1982: 142) stated that the average fixation duration for skilled readers is about 200 to 250 ms +/ - 20 ms during silent reading. Moreover, a fair percentage of fixations exceed 450 ms (Rayner 1983: 117). At the turn of this century, it was generally acknowledged that fixations usually last between 200 ms and 250 ms (Rayner *et al*. 2006: 242). However, in some cases, it can range from around 50 ms to 500 ms (Keating 2013: 72).

Initially, Rayner (1998: 373) stated that the average fixation lasts for 225 ms during silent reading, 275 ms during RA, 400 ms during reading and typing simultaneously, 275 ms during visual search, and 330 ms during scene perception. Rayner (2009: 1460) then revisited the mean fixation duration in different visual tasks and published the range of mean fixation durations in silent reading (225-250 ms), RA (275-325 ms), scene perception (260-330 ms), and visual search (180-275 ms). Moreover, Wolverton and Zola (1983: 47) found that "visual information is noticed throughout the fixation, but […] a critical point of movement decision is made around 140 ms prior to the onset of the fixation". In other words, the decision to cast a fixation and where to place such a fixation have already been made around 140 ms before this fixation takes place.

In general, not all the words in a text receive fixations (Rayner *et al*. 2006: 243). Starr and Rayner (2001) stated that content words are usually fixated upon 85% of the time while function words are fixated upon 35% of the time (ibid.: 158). According to Rayner and McConkie (1976), words that are seven to eight letters long were fixated 95% of the time, whereas shorter words that were five letters long were only fixated 64% of the time. Furthermore, Rayner (1983: 117) found out that "the probability of fixating on a shorter word decreases as word length decreases". It has been reported that the fixation time on words could be influenced by textual circumstances such as low-frequency, misspellings, ambiguity (McConkie and Yang 2003), as well as word length and familiarity (Pickering *et al*. 2004). Long, ambiguous, and difficult words tend to receive longer and more fixations than shorter, easier, and highly predictable words (Liversedge and Findlay 2000; Rayner 1998; Morris and Williams 2004; Sjørup 2013).

Many eye-movement indicators have been fixation-based due to the strong link between cognitive effort and the location of one's fixations (Rayner 1998). Some of these fixation-based indicators are position measures, some of them are duration measures, and some of them are numerosity measures. Among them, total fixation duration (TFD), fixation count (FC), and mean fixation duration (MFD) are the most common (e.g., Pavlović and Jensen 2009; Hvelplund 2014).

TFD is a cumulative measure of fixation durations (Walker 2018), calculated by "combining fixation duration and number of fixations" (Hogaboam 1983: 310). It is said to be "the most used measure in eye-tracking research" (Holmqvist *et al*. 2011: 377). Also called *dwell time*, it is similar to the concept of gaze duration and they are often conflated (Holmqvist *et al*. 2011: 5). Hence, the present study uses the term *fixation duration* to avoid confusion. Many researchers have acknowledged that an increase in the total fixation duration is indicative of a greater cognitive effort devoted to the task (e.g., Jessen, Sjørup, and Balling 2009: 322; Saldanha and O'Brien 2014: 144).

Similarly, a longer TFD and a larger number of fixations are generally considered as an indication of cognitive load associated with an increase in cognitive demands (Just and Carpenter 1980; Rayner 1998; Holmqvist *et al*. 2011). Also, the number of fixations per area unit could be used as "an operational definition of reading depth" (Holmqvist *et al*. 2011: 412). Both total fixation duration and fixation count have been applied by Korpal (2012) to indicate cognitive effort in an eye-tracking based STR research.

FC and MFD seem to have a rather complicated relationship with each other. Researchers using eye-tracking data in reading-process research may assume that the *average fixation duration* and the *fixation count* both index cognitive effort. However, it should be noted that a larger number of fixations and a longer average fixation duration are contradictory to each other. When the trial duration is fixed, there should be a clear negative correlation between the number of the fixations and the mean duration of fixations: the more fixations, the lower the MFD, and vice versa (Holmqvist *et al*. 2011: 412). Therefore, readers may substitute the number of fixations for the MFD, or the other way around (Hogaboam 1983: 310). Which do they tend to substitute for the other? This question has not been addressed or explored in detail, but some studies have shown that subjects can sometimes show an unconscious preference. For instance, in an eye-tracking based study of Sharmin *et al*. (2008), translation students cast a larger number of fixations on the more complex ST than the relatively easier ST, but their total fixation durations were almost the same (2008: 44). Also, the average MFD did not have a significant difference in the two different reading tasks and an STR task carried out by Jakobsen and Jensen (2008). Hence, under these circumstances, the subjects taking part in the experiments might have substituted the length of the mean fixation duration for the number of fixations.

Besides the aforementioned three indicators, there are many other fixation-based indicators. Some researchers (e.g., Rayner 2009) have developed alternative fixation-based measures to deal with specific research questions, such as *first-fixation duration* (the duration of the first fixation on a specific word), *single-fixation duration* (the duration of the only fixation on a specific word), and *gaze duration* (the sum of first-pass fixations on a given word) (Rayner 2009: 1461). Thereafter, Schaeffer *et al*. (2016) rebranded two of them as the First Fixation Duration (FFDur) and the First Pass gaze

Duration on ST words (FPDurS).

4.2.3 Saccade-based Eye-tracking Indicators

Saccades are very rapid eye movements and their purpose is "to place the foveal region on that part of the text to be processed next" (Rayner 1995: 5) during reading. In other words, saccadic eye movements serve to bring the fovea of the eye, which is the point of the highest acuity, from one spot to another (Gibson and Levin 1975: 351). The electrical stimulation of one's midbrain reticular formation is responsible for producing saccadic eye movements and the frontal eye field generates and controls saccadic activities (Breitmeyer 1983: 12-14). The amplitude and direction of the saccades in the visual field are guided by the schemata (Breitmeyer 1983: 3).

Downing and Leong (1982: 143) stated that saccades are rapid conjugate movements that last from 30 to 120 ms. Subsequently, the estimated length was narrowed down to 20 to 40 ms by Rayner and Pollatsek (1989: 113). The saccadic amplitude was reported to typically range from 7 to 9 letters in silent reading and 6 to 7 letters in RA (Rayner 2009: 1460). In 2011, Holmqvist *et al*. (2011: 23) recalculated the estimated length saccades and stated that they typically last from 30 ms to 80 ms.

Pure saccades are not equal to saccadic eye movements in reading. According to McConkie (1983: 70), saccadic eye movements in reading include *forward movements*, *regressions*, *return sweeps*, and *corrective movements*. Regressions, for example, are backward, regressive saccade movements that happen when the readers look back to previous words or lines (Rayner 1998: 373; Starr and Rayner 2001: 157). It should not be confused with return sweeps, which are the "right-to-left saccades from the end of one line to the beginning of the next" (Rayner 2009: 1460). The return sweeps are "not usually counted as regressions because they are moving the reader forward through the text" (Rayner and Pollatsek 1989: 114). According to Keating (2013: 71), 85% to 90% of mature readers' saccade is progressive, and the remaining 10% to 15% is regressive. The majority of regressions fall onto the immediately preceding word, but larger-range regressions might happen when comprehension was not satisfying (Rayner 2009: 1460).

Some researchers argued that the extraction of the visual information from the text solely takes place during fixations (e.g., Wolverton and Zola 1983). Wolverton and Zola (1983) believed that this is because the eyes are moving so fast during saccades that the words would "largely be a smear and thus highly unintelligible" or the brain "sends out a signal to the visual system to ignore (or attenuate) all input from the eyes until the saccade is over" (Rayner and Pollatsek 1989: 122). Due to this kind of saccadic suppression, many researchers believed that "the ability to detect a stimulus is attenuated during saccades" (Wolverton and Zola 1983: 43). In other words, vision during saccadic eye movements is possible but reduced (Volkmann 1962). While the aforementioned researchers argued that a subject's vision is suppressed during saccades, other researchers reported in a few special cases when new information could be acquired (e.g., Campbell and Wurtz 1979). However, Rayner and Pollatsek (1989: 122) argued that even if there is any visual information registered by the sensory system during saccades, it is of little practical importance. In summary, under ordinary circumstances, a reader cannot acquire new information during a saccade (Rayner 2009: 1458). This, however, does not mean that information processing is not occurring during saccades, as Sjørup (2013: 82) argued that it is possible that existing information is still processed during saccades.

Saccade-based indicators have been undeservedly undervalued. Since "virtually all the information is extracted during fixations" (Rayner and Pollatsek 1989: 123), the fixation-based indicators have been used extensively in the eye tracking-based studies. Nevertheless, saccades indicate the changes in one's gaze points. One's covert attention shifts with the fixations, but saccades are how the attention manages to shift. Indeed, there can be shifts of attention independent of saccades. When a saccade is executed, however, a shift of attention is bound to happen. Hence, as saccades exist to "control the flow of information extraction" (Rayner and Pollatsek 1989: 123), the saccade-based indicators should not be ignored. McConkie (1983: 71) argued that since no correlation was found "between the duration of fixations and the lengths of saccades preceding or following them", it shows readers have "independent control of fixation durations and saccade extents". Breitmeyer (1983: 3) also suggested that the control sensory of saccades "is affected by cognitive and by peripheral sensory processes". This, somewhat, supports the using the fixation-related measurements and the saccade-related measurements separately as independent indicators.

Saccade duration has not been commonly counted as a part of visual attention because it is believed that one's vision during the saccade is not very accurate (Gibson and Levin 1975: 352). Holmqvist *et al.* (2011: 23) stated that one could be considered *blind* during saccades. Therefore, eye-tracking studies applying saccade-based indicators are relatively rare. Instead of being used as an indicator of cognitive events, saccade duration has been more commonly studied as a research subject (e.g., Downing and Leong 1982; Rayner and Pollatsek 1989).

Nevertheless, there are still some saccade-based indicators, including saccade length and saccade rate. To be specific, since the average saccade length reflects the region of perceptibility (O'Regan 1979), an increase in the saccade length reflects a more efficient use of peripheral visual information. Hence, a decrease in saccade length is seen to be reflecting "the reduced capacity in short-term memory for holding the incoming information" (Ehrlich 1983a: 262). Another saccade-based measure is called *Saccade Rate*, which is the number of saccades divided by the duration of a trial in seconds, calculated as the number of saccades per second (s^{-1}) (Holmqvist *et al*. 2011: 404). According to Rayner's (1978) estimation, saccades take up about 10% of the total task time (TTT) during silent reading. As the task difficulty and the fatigue level increases, the saccade rate decreases (Van Orden *et al*. 2000; Nakayama, Takahashi, and Shimizu 2002; Pan *et al*. 2004).

When using saccade-based indicators, researchers need to consider a variety of conditions as saccades are highly subjective to the reading task and material. Since "both visual and textual variables influence the size of saccade" (Rayner and Pollatsek 1989: 167), local features such as the space in the particular writing system and the space between lines might have an impact on saccade patterns. Abrams and Zuber (1972) computed the probability of a fixation landing in the blank spaces of several characters' length. The probability was nil. The result demonstrated the effect of the space between words on one's saccade control and excluded the possibility of using saccade-based indicators to compare reading/translating an alphabetic language with a logographic language. Therefore, it would be problematic to compare reading-related tasks in different languages, such as English and Chinese, with the help of saccade-based indicators because English words are separated by spaces whereas Chinese characters are not.

4.2.4 Combined Indicators

Although “somehow the brain is able to smooth out the discrete inputs from each fixation and create a feeling of a continuous coherent perceptual world” (Rayner and Pollatsek 1989: 141), the notion that the eyes and the mind sweep continuously across the text, pause or regress when difficulties were encountered is an illusion (Rayner and Pollatsek 1989: 113). As mentioned at the beginning of Section 4.2, eye movements are a combination of fixations and saccades. It was found that the fixation durations and saccade lengths vary in different reading tasks (Rayner *et al*. 2007), but they do not correlate with each other (Castelhano and Henderson 2008). After all, fixations are conscious or voluntary movements, whereas saccades are subconscious movements (Clement and Sørensen 2008: 150). According to Rayner (1983: 115), “fixation duration and saccade length are influenced by the cognitive processing state of the reader and by the characteristics of the visual patterns in the text”.

Since there is a very close and complicated relationship between these two most basic and yet essential gaze behaviours, combined indicators that consist of both countable entities are often applied to analyse the eye movements and cognitive processing during visual activities. For example, TTT, gaze samples, and pure gaze activities (also called the sum of Fixation and Saccade Durations, FSD) are combined indicators used in some previous eye-tracking based translation studies (e.g., Liu, Zheng, and Zhou 2018). Among these examples, TTT is used most frequently because task complexity has a direct impact on the amount of time spent completing the task (Shreve and Diamond 1997). The use of TTT in previous studies (e.g., Pavlović and Jensen 2009) is based on O’Brien’s (2008: 85) argument that “processing speed is a good measurement for cognitive effort on the basis that difficult tasks generally take longer than easier tasks”. Essentially, these combined indicators mentioned above are on a hierarchy of the amount of data they consist of: the pure gaze activities are the sum of pure fixations and saccade durations; the gaze samples contain pure gaze activities and noise; the sum of gaze samples and non-gaze samples equals the total task time. In summary, these indicators are all dependent variables that can be used to study mental processing.

Chapter 5. Research Design and Data Acquisition

This chapter begins with the research design, including subjects, stimuli, experimental tasks, piloting, and formal experimental procedures, which is described in Section 5.1. Methodological considerations concerning the apparatus, the fixation filter, *Segment* and *Scene*, *Area of Interest* (AOI), and EVS coding and calculation are explained in Section 5.2. Finally, the eye-tracking data acquisition, in particular data quality assessment, is illustrated in Section 5.3.

5.1 Research Design

5.1.1 Subjects

According to Göpferich *et al*. (2008), the average number of subjects in eye-tracking based translation studies is 12. In the past ten years, however, this number has increased to some extent. The present experiment recruited three groups of subjects including a large number of subjects and trials, with the hope of gaining better *statistical power* (Holmqvist *et al*. 2011: 86).

Excluding two subjects who did not pass calibration, Group A comprised 31 (27 female, 4 male) voluntary and consenting Mandarin-Chinese speakers who were students in MA Translation Studies at Durham University (average age = 24.29, *SD* = 1.87). They had the same language background (Mandarin Chinese as L1, English as L2) and similar English proficiency (Mean IELTS score = 7.31, *SD* = 0.33). None of the subjects had any professional translation or interpreting experience. The subjects were recruited as volunteers for four tasks including Reading Aloud in Chinese (RA-C), Chinese-English STR (STR C-E), Reading Aloud in English (RA-E), and English-Chinese STR (STR E-C). Each subject's experiment lasted around 50 minutes. The experiment took place in their final academic term, two to three months before they completed their MA degree. The subjects were assigned the codes A01-A31.

Group B consisted of 36 (15 female, 21 male) voluntary and consenting native speakers of English studying at Durham University (average age = 22.89, *SD* = 1.39). All were monolingual English speakers born and raised in the UK. Subjects were recruited as volunteers for the task Reading Aloud in

English (RA-E) experiment. Each subject's experiment lasted for around 10 minutes. The subjects were labelled B01-B36.

Group C comprised 21 (11 female, 10 male) voluntary and consenting native speakers of German studying at Durham University (average age = 23.05, *SD* = 1.88). They were all Erasmus or MA students who had the same language background (German as L1, English as L2). The average score of those who had taken the IELTS examination was 7.41 (*SD* = 0.38). None of the subjects had any professional translation or interpreting experience. These subjects were recruited as volunteers for the tasks Reading Aloud in English (RA-E) and English-German STR (STR E-G). Each subject's experiment lasted for around 30 minutes. The subjects were assigned the codes C01-C21.

All the subjects signed a consent form (see Appendix 3) indicating their willingness of participation in the experiment, which was approved by the Research Ethics Committee of the School of Modern Languages and Cultures at Durham University.

Recruiting subjects was somewhat problematic. The individuals' eyesight and glasses had to be taken into consideration. Jensen (2008: 167) showed that there can be issues with glasses and contact lenses during eye-tracking experiments, but he did not choose to filter out subjects based on the use of glasses or contact lenses in the first instance. The present study took the same course of action. Instead of excluding subjects wearing either glasses or contact lenses beforehand, the present study completed the experiments while the subjects wore them and then assessed the quality of the eye-tracking data based on a set of criteria. There is a chance that wearing visual aids might sacrifice the quality of the gaze data recorded by the eye-tracking equipment, but it is also possible that the data quality would be unaltered. For the same reason, subjects with hooded eyelids or/and downward pointing eyelashes were not eliminated before the experiments.

Even though glasses and contact lenses are not considered to be a problem when collecting eye-tracking data, factors such as narrow-framed eyeglasses, bifocal lenses, long eyelashes, and heavy mascara can obstruct the participant's eyes and, as a result, impact the data quality (Hvelplund 2014: 207). Therefore, the researcher contacted the subjects one day in advance of the experiments, reminding them to wear contact lenses or glasses if necessary, to not wear heavy eye makeup, and to avoid alcohol and caffeinated drinks before the experiment. To this end, all subjects had normal or corrected-to-normal vision.

The present study recognises that there were more female subjects than male subjects. Other studies have faced the same situation, such as Göpferich (2010), who had 11 female and 1 male subjects in a test group. This lack of balance was not an issue because I believe that gender difference does not have a significant impact on one's pupil response to cognitive workload (e.g., Beatty 1982; O'Brien 2008). In addition, achieving gender balance was extremely difficult, especially in Group A, as the vast majority of the MA translation students were female.

5.1.2 Stimuli

Different types of materials can be used when designing eye-tracking experiments that investigate reading and translation processes. Many studies have used non-linear reading materials, such as a list of words and phrases in their experiments (Hyönä, Radach and Deubel 2003: 314). In recent years, however, empirical studies on translation have been using longer and linear text as STs (e.g., Jakobsen and Jensen 2008; Hvelplund 2011; Zheng and Zhou 2018). Reading studies often employ reading materials designed specifically for research purposes, whereas translation studies tend to use authentic STs. This choice to use authentic STs stems from Deignan's work that pointed out that using constructed STs in experiments "may be forcing participants to tackle problems that are not faced in normal discourse" (2005: 117). O'Brien (2009: 261) argued that the texts used should be 200-300 words, pointing out that the subjects could get tired and lose motivation if the experiment is too long. Now, researchers usually select authentic texts between 100-200 words for their written translation or STR experiments (e.g., Dragsted 2010; Shreve *et al*. 2010). While 100-200 words may seem short, even short texts can yield a significant amount of data (Saldanha and O'Brien 2014: 140).

There were two stimuli for this research: Material A that consisted of 10 sentences, 200 English words (see Appendix 1) and Material B that consisted of 12 sentences, 426 Chinese characters (see Appendix 1). The average number of words in a sentence in Material A is 20 (*SD* = 14.03). The average number of words in a sentence in Material B is 35.42 (*SD* = 22.13). The word counts of all the sentences in both Material A and Material B are within 2.5 standard deviations of the average. Therefore, although the word count fluctuates from sentence to sentence, none of the sentences are outliers according to researchers working with statistical computing (e.g., Crawley 2002). This kind of variation in sentence length is typical when using authentic texts.

Material A was used as the reading material and the ST for the tasks of Reading Aloud in English (RA-E), English-Chinese STR (STR E-C), and English-German STR (STR E-G). In Zheng and Xiang's studies (2013, 2014), they employed a piece of original English text as the ST for their experiments with the intention of "imitat[ing] a real-life translation scenario as closely as possible" (2013: 164). The particular excerpt from US President Bill Clinton's farewell address was chosen to be the STR ST in Zheng and Xiang's (2013) studies. The same text was used by Zheng and Zhou (2018) to be consistent with the previous studies (Zheng and Xiang 2013, 2014). The only difference was that one long sentence in Zheng and Xiang's (2013, 2014) ST was removed by Zheng and Zhou (2018), so that the content on every page was delivered by Clinton in approximately 20 seconds. This ST extract (Material A) from Bill Clinton's farewell address (see Appendix 1) was employed as the reading material for Reading Aloud in English (RA-E), as well as the ST for English-Chinese STR (STR E-C) and English-German STR (STR E-G) in the present research.

The text statistics and readability indices (the five US grade level indices[1]) of Material A were calculated by Editcentral.com. The indicators are shown in Table 5.1a.

Table 5.1a: The Text Statistics and Readability Indices of Material A

Material A	Text Statistics	Readability Indices
	Word Count 200	Flesch Kincaid Reading Ease 59.6
	Complex Words 22	Flesch Kincaid Grade Level 9.9
	Sentences 10	Gunning Fog Score 11.6
	Words per Sentence 20	SMOG Index 8.7
	Average Syllables per Word 1.5	Coleman Liau Index 10.7
	Percentage of Complex Words 11%	Automated Readability Index 9.8

During the experiment, the text was evenly distributed across four slides. The original speech content on each page was delivered by Clinton in approximately 20 seconds. Subjects in the formal experiment were given 30 seconds to read the content on each page (120 seconds in total) in the RA-E task and 50 seconds to sight translate the content on each page (200 seconds in total) in the STR tasks. Subjects were asked to look at the page number on the bottom once they finished reading and translating each page. When the 30 and 50 seconds ran out, the slide on the screen automatically turned to the next page.

Material B was used as the reading material and the ST for Reading Aloud in Chinese (RA-C) and Chinese-English STR (STR C-E). The Chinese text was a translation from another excerpt from Bill Clinton's farewell address (see Appendix 1). It was published on the Chinese online news website Sohu News (http://news.sohu.com) in Chinese. The ST consists of 426 Chinese characters in total, almost evenly presented on four PowerPoint slides. The original speech content (in English) on each page was delivered by Clinton in approximately 20 seconds. Same time limits were used for Reading Aloud in English and English-Chinese STR, these two tasks gave subjects 30 seconds to read the content on each page (120 seconds in total) and 50 seconds to sight

---

[1] The five US grade level indices, namely Automated Readability Index (ARI), Flesch-Kincaid, Coleman-Liau, Gunning fog and SMOG, are designed to indicate comprehension difficulty and gauge the understandability of an English text. The formula of each index is as follows (obtained from Editcentral.com):
ARI= 4.71*characters/words+0.5*words/sentences-21.43;
Flesch-Kincaid =4.71*characters/words+0.5*words/sentences-21.43;
Coleman-Liau=5.89*characters/words-0.3*sentences/(100*words)-15.8;
Gunning fog =0.4*(words/sentences+100*((words >= 3 syllables)/words));
SMOG=square root of (((words >= 3 syllables)/sentence)*30) + 3.

translate the content on each page (200 seconds in total). Subjects were asked to look at the page number once they finished reading and translating each page. When the time set for each page ran out, the slide on the screen automatically turned to the next page.

Although there are “no standard guidelines for font sizes” (Saldanha and O’Brien 2014: 136), large font sizes are generally preferred because the imprecise registration of fixation data could be “off-set by increasing the line spacing and/or the font size” (Dragsted and Hansen 2009: 591). To be specific, translation studies that use an eye-tracking method typically opt for font sizes between 16 and 20 (Hvelplund 2014: 208). In addition, the choice of font size is practically relevant to the research design when the analysis is at the sentence level or the lexis level (ibid.). Following Zheng and Zhou (2018), the present study used Times New Roman font at size 16 with a line space of 8. The text is black on a white background. Based on Saldanha and O’Brien’s suggestion (2014: 136), the text did not extend to the very right or left of the screen as fixations on the edges might have been lost.

5.1.3 Experimental Tasks

The subjects in Group A were asked to complete four tasks: Reading Aloud in Chinese (RA-C), Chinese-English STR (STR C-E), Reading Aloud in English (RA-E), and English-Chinese STR (STR E-C). The subjects in Group C were asked to complete two tasks: Reading Aloud in English (RA-E) and English-German STR (STR E-G). Subjects in Group B were asked to perform only one task: Reading Aloud in English (RA-E). The research experiment randomised the presentation of the tasks based on Saldanha and O’Brien’s suggestion (2014: 117). The presentation order of the materials, therefore, was interchanged among the subjects in Group A. Half of the Chinese subjects was asked to complete Reading Aloud in English and English-Chinese STR first and the others were asked to complete Reading Aloud in Chinese and Chinese-English STR first.

The present study applied the “Prepared STR with a time limit” (Zheng and Xiang 2013; Zheng and Zhou 2018). It is a STR variety that has been widely seen in the real-life practice, such as in court trials, community interpreting, and business negotiations. Compared to written translation tasks performed with no time pressure, this type of STR involves a more intense integration of the input and output, thus demanding a more condensed coordination effort in a given amount of time. Although STR translators can control the speed of delivery, they should know that "listeners expect the product to be presented in fluent, connected speech" (Jakobsen and Jensen 2008: 116). As a result, the time pressure makes this modality significantly different from written translation, but similar to simultaneous interpreting.

Compared with silent reading for comprehension, RA requires additional effort (Jensen 2008: 160). The oral production during RA is a linear, non-hierarchical presentation of the segments shown in the reading material, whereas the output of STR consists of the equivalences of the ST units that are frequently in an altered sequence. According to Jensen (2008: 160), RA is "the best type of normal, linear reading activity which allows for simultaneous monitoring of the comprehension process and the associated gaze behaviour". Also, Jakobsen and Jensen (2008: 121) stated that they should have included a RA task to "know more clearly how much additional eye movement was caused by a concurrent language production activity that did not involve translation" in their study on eye movement behaviour across four different types of reading tasks.

One might argue that since the RA and its corresponding STR tasks share the same piece of reading material, *priming* will take place in the STR tasks. Priming refers to "a psychological effect that affects language in response to stimuli so that the prior encountered element is repeated or processed faster" (Bangalore *et al*. 2016: 217). The present study, however, chose to design the experiment in this way because it was vital to maintain the comparability of the reading materials and the STs. In addition, this priming effect influenced all of the subjects, rather than just some of the students. Therefore, the priming effect coming from the RA experiment could be viewed as part of the preparation for the STR.

Before each of the formal tasks, there was a warm-up exercise to help subjects get used to the experiment setting and the task requirements. Material C (see Appendix 1) was another excerpt from Bill Clinton's farewell address and also came from the published translation of Bill Clinton's farewell address on Sohu News. It consisted of 2 slides. Subjects were given 30 seconds to read the content aloud and 50 seconds to sight translate the text. The warm-up exercises helped the subjects understand the experiment settings such as the time limitation, font, size of the words on the screen, and the experiment procedures.

Stansfield's (2008: 15) sight translation guidelines propose giving sight translators the chance to familiarise themselves with the content and provide necessary glossaries to ensure a smooth interpretation. The preparation process proved to be necessary in the pilot study as the subjects reported that they were unable to complete the STR task without preparation. Hence, before the formal English-Chinese STR task and the English-German STR task, subjects were given three minutes to prepare for the experiment with a copy of the ST and relevant glossary (see Appendix 2). The glossary list was published by Zheng and Zhou (2018). Before the formal Chinese-English STR task, the subjects were also given three minutes to prepare for the experiment with a copy of the ST. Subjects received a reminder when there were 50 seconds left, which is also how much time they were given to sight translate each page of the ST. The ST and the glossary were removed after three minutes.

5.1.4 Piloting

An exploratory pilot study was carried out to investigate EVS in sight translating from English to Chinese that recruited 19 subjects who were students at Durham University. They sight translated the ST which was presented on four slides, with each slide being displayed for 40 seconds. The STR processes were registered by a Tobii TX300 eye-tracker and an audio recorder pre-installed with the eye-tracking programme. Then, the retrieved eye-tracking and audio data were analysed by Tobii Studio and the Audacity software. The pilot study proved the general existence of reading ahead activity during STR, laying the foundation for the research on EVS during RA and STR. The presence of EVS during STR also provided a potential answer for a debate in translation process research: is translation a sequential process or a series of comprehension and production activities occurring in parallel? (Carl and Dragsted 2012; Balling *et al*. 2014). The results from the pilot study showed that during STR, "source and target related processes are tightly intertwined" (Schaeffer *et al*. 2016: 187). Besides, the time of reading ahead into ME was greater than that of reading ahead beyond ME in most cases, indicating that the length or size of EVS could fluctuate depending on the local processing difficulty caused by the ST.

Following the pilot study, the researcher made the following adjustments based on the feedback:

1. The font and the size of the words remained the same, but the slides were enlarged by 200%. This change made the ST more visible to subjects who had minor near-sightedness but did not have glasses;
2. To reduce the number of saccades moving from the end of the upper lines to the head of the lower lines, the direction of the slides was changed from vertical to horizontal. As a result, the number of lines reduced from 20 to 16 in Material A. The slides for Material B were made horizontal for the same reason;
3. The length of preparation time was not adjusted (3 minutes), but the time for sight translating materials on each slide was extended from 40 seconds to 50 seconds. In the pilot study, only 15 subjects were able to complete the translation within 40 seconds. The total STR time was set to be 50×4 seconds based on Jakobsen and Jensen's (2008: 108) experiment, in which the average amount of the time the subjects needed to sight translate the approximately 200-word English passage was 204 seconds;
4. One more criterion named the Saccade as a Percentage of Gaze Sample (SPG) was added to the quality assessment process. It was based on the Gaze Sample to Fixation Percentage (GSF/ GFP) applied by Hvelplund's (2011) study. It was also tested in the pilot study.

Subsequently, two subjects participated in a small-scale pre-test to further test the adjusted experiment setup. Their feedback showed that the preparation time was sufficient and the time given to read and sight translate the content on each page was adequate.

5.1.5 Experimental Procedures

An individual session was scheduled for each subject. The researcher briefly explained the experimental procedures before each formal experiment.

Subjects were asked to complete the tasks on their own. There were no online or offline translation aids available during the STR process. Subjects were also informed about the time limit and the requirement that they aim to produce high-quality STs before the STR tasks. To avoid extreme gaze angles, the researcher told the subjects that they could lift or drop their chin slightly when reading the first and the last line.

In a previous eye-tracking study conducted by Schmaltz *et al.* (2016: 244), subjects were told to "keep their eyes on the monitor as much as possible", but they could "move freely" during the experiment. However, in the present eye-tracking study, all the subjects were able to adjust their sitting position before each of the tasks and were required to hold the position during that task. This instruction might raise a challenge for ecological validity, but it ensures that an adequate amount of eye-tracking data can be successfully registered.

The eye-tracker was calibrated at the beginning of all the warm-up exercises and all the formal tasks. The purpose of conducting a *calibration* is "to gauge how well the eye-tracker can capture an individual's eye movements" (O'Brien 2006: 188). This eye-tracking experiment employed the regular 5-points-calibration at a medium speed, which consisted of five points presented on the screen, fixated and sampled one at a time. When a poor calibration was detected at the beginning of a trial, a recalibration was undertaken.

The experiment procedures carried out with Group A were:

1. The researcher briefed the subjects on the experiment procedures and the origin of the reading materials;
2. The order of tasks requiring L1 or L2 processing was randomised. The researcher assigned a warm-up exercise of Reading Aloud in English/Chinese for the subjects to get used to the experiment setup and the environment;
3. The subjects were asked to complete the Reading Aloud in English/Chinese task;
4. The subjects completed the warm-up exercise for the English-Chinese/Chinese-English STR task;
5. The subjects were given three minutes to prepare for the STR

experiment with a copy of the ST. If they were doing the English-Chinese STR task, they were also given a relevant glossary list in this preparation stage. They received a reminder when there were 90 seconds left as it was half of the preparation time and a reminder when there were 50 seconds left as this was the amount of time allocated to sight translating each page of the ST;

6. The subjects completed the English-Chinese/Chinese-English STR task;
7. Finally, both the researcher and the subjects signed the consent form.

The experiment procedures carried out with Group B were:

1. The researcher briefed the subjects on the experiment procedures and the origin of the reading material;
2. The researcher assigned a warm-up exercise for subjects to get used to the experiment setup and the environment;
3. The subjects were asked to complete the task Reading Aloud in English;
4. Finally, both the researcher and the subjects signed the consent form.

The experiment procedures carried with Group C were:

1. The researcher briefed the subjects on the experiment procedures and the origin of the reading material;
2. The researcher assigned a warm-up exercise of Reading Aloud in English for the subjects to get used to the experiment setup and the environment;
3. The subjects were asked to complete the formal Reading Aloud in English task;
4. The subjects started the warm-up exercise of English-German STR;
5. The subjects were given three minutes to prepare for the experiment with a copy of the ST and a relevant glossary list. They received a reminder when there were 90 seconds and 50 seconds left;
6. The subjects completed the English-German STR task;
7. Finally, both the researcher and the subjects signed the consent form.

5.2 Experiment Settings

5.2.1 Apparatus

The series of reading and STR experiments were carried out at the School of Modern Languages and Cultures, Durham University, from April to October 2016. To preserve the ecological validity and to best imitate an authentic STR working scenario, a Tobii TX300 eye-tracker was used in the present research to record binocular eye movements with a sampling rate of 300 Hz. It is a remote video-based eye-tracking system integrated with a removable 23'' TFT monitor at 1280×1024 pixels. Subjects' oral outputs were recorded by The Logitech 980186-0403 USB noise-cancelling desktop-microphone with a frequency range of 100 Hertz (Hz) - 16 Kilohertz (KHz). The experiments did not employ any immobilising apparatus, so the subjects were instructed to minimise their body and head movements.

The orientation of the monitor was landscape with the aspect ratio at 16:9. The screen refresh rate was 60 Hz. The display colours were 16.7 M (Hi-FRC). The vertical sync frequency and the horizontal sync frequency were 49-75 Hz and 54.2-83.8 Hz respectively. For this 300Hz eye-tracking system, there are 300 samples recorded per second. According to Holmqvist *et al.* (2011: 31), a 50 Hz eye-tracker is a relatively slow system. The higher the sampling frequency, the more precise the measure of the onset and the offset of different eye movements is.

All experiments were carried out at the same location. The window-less enclosed laboratory was equipped with an overhead fluorescent light to keep the lighting in the eye-tracking laboratory consistent, which meant the eye movements and the pupil size were less likely to be influenced by any variation in light intensity. The laboratory was not soundproof as soundproofing the room was beyond the means of this translation research. The laboratory is located away from any loud noises and the experiments were not scheduled during busy hours to reduce possible interferences. These efforts were made to reduce the chance of distractions, sudden pupil dilation and constriction.

In general, it is typically recommended that the subjects sit around 60-80 centimetres (cm) away from the monitor (Hvelplund 2014: 209), but researchers have the freedom to set the viewing distance between their subjects and the monitor based on their experiment settings. For example, Sharmin *et al*. (2008) set the viewing distance between 50 cm and 60 cm; Jakobsen and Jensen (2008) and Korpal (2012) set it around 60 cm; Schmaltz *et al*. (2016) set it around 55 cm; Walker (2018) set it between 60 cm and 65 cm. Based on these studies and the pilot experiment, the viewing distance between each subject and the monitor was set at between 60 cm and 65 cm.

5.2.2 Fixation Filter

The Tobii TX300 eye-tracker collects raw eye movement data points every 3.3 to 33 ms. The fixation filters are responsible for calculating the fixation data, such as fixation count, duration, and location[2]. The word *filter* used in the eye tracking analysis software and particularly in the Tobii software, refers to "the complete set of steps available to identify fixations within the raw data" (Olsen 2012: 4).

With the launch of Tobii Studio 2.3 in June 2011, the function of following the eye movements in 3D space was improved (Olsen 2012: 4). According to Tobii Technology, "this ability allows for more sophisticated fixation filters to be developed of which the new Tobii I-VT filter is an example" (ibid.). Of course, there are several other options when it comes to selecting a fixation filter: the Tobii Fixation Filter, the ClearView Fixation Filter, and the Raw Data Filter.

The standardisation of the settings should be achieved through the selection of a filter to maintain the validity and reliability of the eye-tracking data (Alves *et al.* 2009: 274). Since the present research is closely related to Zheng and Zhou's (2018) study, all the experiments in this study used the same filter. This is, however, not the only reason for choosing the I-VT Filter. This filter was chosen in the first instance because it enables the classification of eye-tracking data as fixations, saccades, or unclassified accurately.

The general idea behind an I-VT filter is the *Velocity-Threshold Identification* algorithm, which classifies the eye movements according to the velocity of the directional shifts of the eyes by the visual degrees per second (°/s) (Olsen 2012: 5). According to the User Guide of Tobii Studio 3.3, "if the

---

[2] Page 51. Tobii Studio 3.3 User Guide.
Available at http://acuity-ets.com/downloads/Tobii%20Studio%203.3%20User%20Guide.pdf

velocity of the eye movement is below a certain threshold the samples are classified as part of a fixation" (52). The default value for the *velocity threshold* parameter was set to 30°/s, which is "sufficient for recordings with various levels of noise" (Olsen 2012: 17) according to Tobii technology.

The default value for the *minimum fixation duration* parameter is set at 60 ms. The max angle between fixations is 0.5 degrees. The default value for the *max time between fixations* parameter (75 ms) is set to be lower than what is commonly reported as a standard duration for short blinks. For the same reason, the default value for the *max gap length* parameter is also set at 75 ms because it is not supposed to fill in normal blinks and gaps when the subject turned away from the eye tracker or when the eye tracker's view of the subject's eyes was obstructed by things like the frame of the glasses (Olsen 2012: 10-18).

The available functions of the Tobii I-VT filter are gap fill-in, eye selection, noise reduction, velocity calculator, I-VT classifier, merge adjacent fixations, and discard short fixations (Olsen 2012: 4). These functions serve to provide accurate fixation classifications collected during eye-tracking experiments and standard reading studies[3]. The filtered data was further analysed to answer specific research questions.

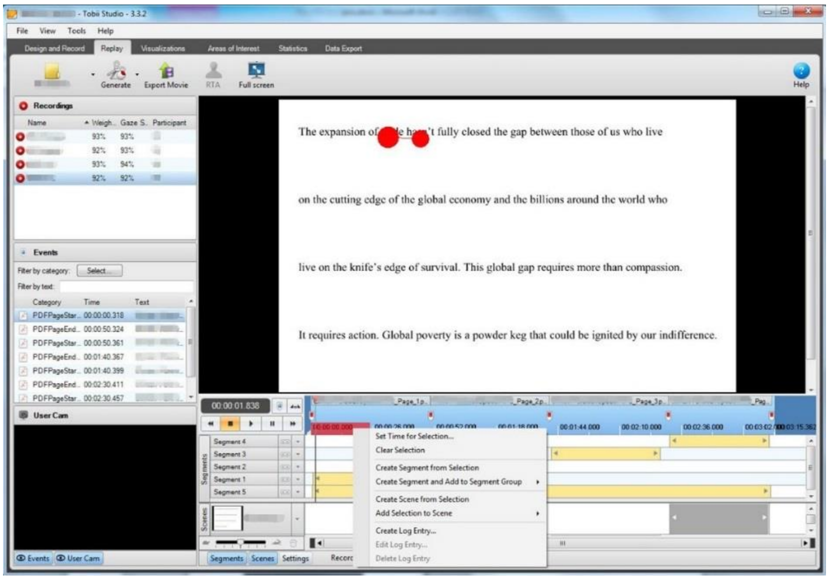
5.2.3 Segment and Scene

A *Segment* is a section of a recording collected by the eye-tracking software application Tobii Studio. A *Scene,* on the other hand, refers to a specific period when the subject's behaviour is under investigation. It can be viewed as a shorter unit within a segment. The scene and segment functions are similar, but scene is more convenient for extracting information about fixations. Researchers can create and set the start and the end of a segment or a scene manually using the timeline marker.

Figure 5.2a shows the Replay function of the Tobii Studio, where both the segments and the scenes were drawn.

---

[3] Page 53-55. Tobii Studio 3.3 User Guide.
Available at http://acuity-ets.com/downloads/Tobii%20Studio%203.3%20User%20Guide.pdf



The expansion of [illegible]le ha[illegible]'t fully closed the gap between those of us who live

on the cutting edge of the global economy and the billions around the world who

live on the knife's edge of survival. This global gap requires more than compassion.

It requires action. Global poverty is a powder keg that could be ignited by our indifference.

Figure 5.2a: Drawing Segments and Scenes in Replay

In this study, segments started from the first fixation cast at the beginning of the passage and ended when the subjects finished reading or sight translating the content on the page. For instance, a Chinese subject had four segments in each task because one segment contained his/her eye-tracking data on one page. After this subject completed the experiment, there were 16 segments in total. Data collected during the segments was then exported to analyse measurements used in Section 5.3, such as unclassified samples and saccade duration and measurements analysed in in Chapter 6, such as pupil diameter.

Scenes were drawn from the first fixation cast at the beginning of the passage on a page to when the subjects finished reading or sight translating the content on the page to analyse measurements discussed in Chapter 6, such as total fixation duration (TFD), mean fixation duration (MFD), and fixation count (FC). Scenes are temporal measures used when we hope to extract certain type of eye-tracking data in a file. However, they do not work without the spatial measures, which is explained in the next section.

5.2.4 Area of Interest (AOI)

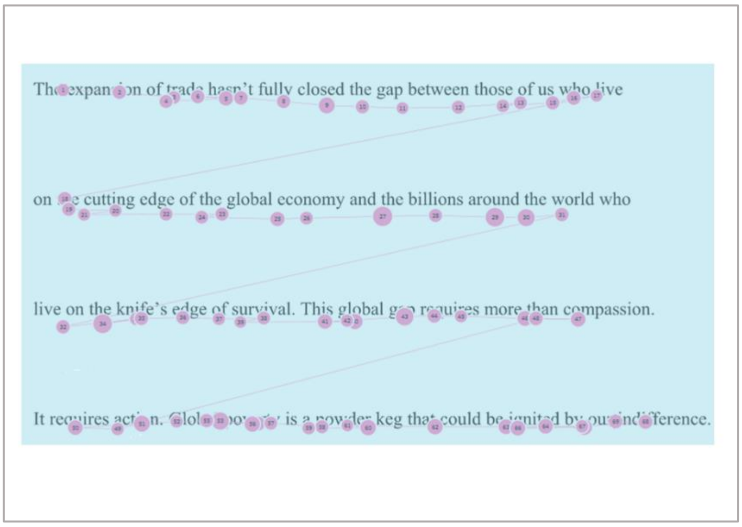
In order to study the gaze behaviour when subjects were viewing a specific text unit on the screen, the researcher often drew this unit as an *Area of Interest* (AOI), then analysed the fixations that fell into this area. The AOIs are also called the Look Zones or the Interest Areas (IAs) depending on which eye-tracker is being used (Holmqvist *et al*. 2011: 5). Furthermore, AOIs are the spatial domain used when we hope to extract certain eye-tracking data. They do not work without using the temporal domain called scene, which is elaborated in the previous section. AOIs are usually rectangular, sometimes polygonal, depending on the layout of the ST under investigation. The AOI editor supplied by the eye-tracking analysis software allows us to draw the spatial dimensions against the background of the stimulus.

Jensen (2008: 159) reminded us that there is “a tendency for the eye to drift down during left to right reading”. Therefore, researchers suggested some solutions to accommodate drift eye movements that fell into the blank space between the lines. Sjørup (2013: 103), for example, specified in her research that “the borders of the AOIs in ClearView were defined to extend halfway into the space to the words on both sides of the individual AOI and also halfway to the lines above and below the line of the AOI”. However, extending the borders to exactly halfway into the space between words and lines ignores the fact that individuals could have different gaze spot mappings.

The present study attempts to address drift eye-movements and at the same time minimise the interference between lines with the following two steps:

1. The line spacing was set at eight times. The width was not typical and the space between lines was relatively large. However, if it were set out using single space or the double spaces, there would be confusion regarding whether a series of fixations was on the upper or the lower line. Having a large space between lines enabled the researcher to draw AOIs that cover the fixations that drifted below or over the written words on the screen;
2. Some subjects’ eye fixations were slightly more consistently above or below the written words as a result of individual reading habits or eye movement traits. In this present research, the researcher chose to manually cut the AOIs for each subject after examining the overall gaze spot characteristics shown in the visualisation page. In short, an individual’s AOIs were personalised based on the visualisation of their reading pattern.

Figure 5.2b is an example of one AOI used to elicit one subject’s total fixation duration, fixation count, and mean fixation duration on the entire ST on one page. These measurements are discussed in Chapter 6.



The expansion of trade hasn't fully closed the gap between those of us who live

on the cutting edge of the global economy and the billions around the world who

live on the knife's edge of survival. This global gap requires more than compassion.

It requires action. Global poverty is a powder keg that could be ignited by our indifference.

Figure 5.2b: An AOI for Eliciting Eye-tracking Data on Page 1

The total fixation duration on the ST here is the sum of the duration of all the fixations that fell into this AOI during the scene covering the reading process. The individual's AOI on this particular page might be slightly different. As mentioned above, some subjects' fixations might be slightly above the first line or slightly below the last line. Therefore, AOIs on the entire text on pages were drawn respectively for different subjects.

5.2.5 EVS Coding and Calculation

Spatial EVS and temporal EVS are the spatial and the temporal aspects of the same phenomenon during reading-speaking processes. As explained in Chapter 3, spatial EVS refers to the amount of text that is within the gap between a fixation and concurrent oral production. Temporal EVS, on the other hand, is the time span between the time when the eyes and the voice go past a specific point in the textual material.

Spatial EVS in RA has been studied in depth using sentences. Sentences are natural segments in a passage, which could be maintained in readers' memory to complete complex semantic interpretations (Slowiaczek 1983: 354). Using sentences for this analysis allows us to account for the natural constraints within and across the sentence boundaries. A marked correlation was discovered between the length of spatial EVS and the position in a sentence: spatial EVS in RA tends to be widest at the beginning of a sentence (Buswell 1921: 221) and narrower at the end of sentences according to both

Quantz (1897) and Buswell (1920). Hence, by calculating temporal EVS at the sentence initials and the sentence terminals, the present research hopes to gain a better understanding of the similarities and differences between spatial and temporal EVS.

Temporal EVS is sometimes measured by the total duration of *off sync fixations*, which are taken to indicate planning or monitoring (Dragsted, Hansen, and Sørensen 2009: 307). EVS calculation in the present study, however, is not limited to the duration of off sync fixations. Instead, the present study adopts "the only really satisfactory way to measure the eye-voice span by actually making a record of the eye movements and relating the eye-movement record to a record of the vocal output" (Rayner and Pollatsek 1989: 181).

Inspired by Laubrock and Kliegl's (2015) way of measuring onset-to-onset temporal EVS, the present research calculates onset-to-onset temporal EVS at sentence boundaries. The first fixation (F_1) landing on the first reading/translation unit and the first fixation (F_x) landing on the last reading/translation unit in a sentence are the two critical fixations used to measure temporal EVS. These two critical positions at sentence boundaries (the first and the last unit in a sentence) are referred to as Sentence Initial (S-I) and Sentence Terminal (S-T). S-I and S-T are definitions in space. Although the researcher is defining sentence boundaries spatially here, temporal EVS in the present study is measured from gaze to speech. The temporal EVS measured at these two positions based on F_1 and F_x, therefore, can be called initial temporal EVS and terminal temporal EVS.

The operational definition of this measure is elaborated as follows: when a subjects' scan path comes to sentence n+1 during RA/STR, a series of fixations and saccades take place in order. Both the first fixation (F_1) at S-I of the current sentence and the first fixation (F_x) at S-T of the current sentence are identified and isolated. To identify F_x at S-T the last reading/translation units in individual sentences are listed below in Table 5.2a.

Table 5.2a The Last Reading/Translation Unit in a Sentence Segment

The Last Reading/ Translation Unit	
English Text	1. the knife's edge of survival
	2. more than compassion
	3. action
	4. our indifference
	5. entangling alliances
	6. from the world
	7. a shared responsibility
	8. all across the world
	9. one America
	10. our common humanity
Chinese Text	1. 一个大家庭
	2. 而努力
	3. 种种分歧
	4. 这个国家
	5. 而奋斗
	6. 还在后面
	7. 不会结束
	8. 更为神圣的契约了
	9. 更为自豪的了
	10. 你们
	11. 你们
	12. 美国

As mentioned in Section 5.1.2, sentences in the reading texts/source texts are of varying length. For a short sentence like *It requires action*, if a subject had to read all three words before he/she began to articulate the translation, then it makes sense to say that the initial and terminal position of this sentence are the same for this subject. However, in the majority of the cases (65.96%), articulation began before subjects looked at *action*. Therefore, looking at *action* as a separate unit at terminal position was justified to some extent in the present study and an independent effect here is expected. Figure 5.2c presents an example of eight individual fixations on one page to calculate temporal EVS, which is examined in detail in Chapter 7.

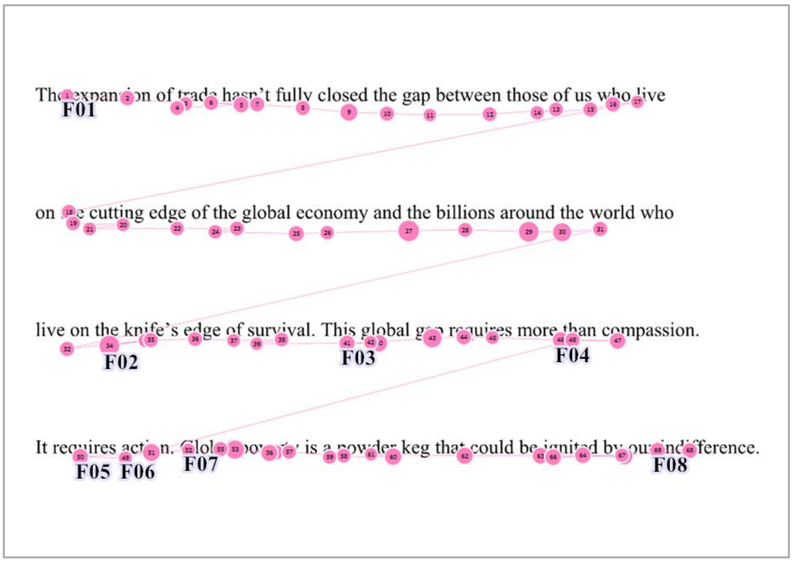
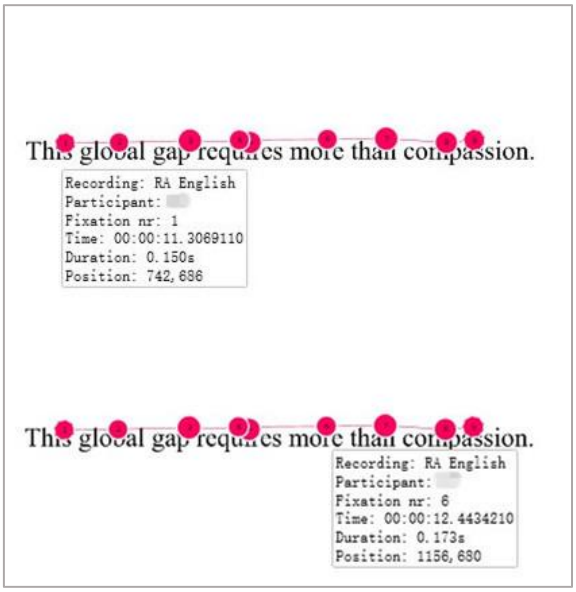


Figure 5.2c: Isolating Fixations to Calculate Temporal EVS

Four of the eight fixations (F1, F3, F5, F7) are the first fixations that fall on the four sentences on this page. Four of the eight fixations (F2, F4, F6, F8) are the first fixations that fall on the last meaning units in the four sentences on this page (the knife's edge of survival; more than compassion; action; our indifference). The onset of these two critical fixations can be noted by splitting the translation logs into time windows (*scenes*) within the function *replay* and drawing individual AOIs to cover the critical fixations with Tobii Studio 3.3.2. To be specific, the places where the F_1 and F_x landed in a sentence are drawn as AOIs. The time of F_1 onset and F_x onset can be obtained using the *statistic* function. However, Tobii Studio offers another option that is more efficient: the onsets of individual fixations can be obtained using *visualisations*.

Figure 5.2d is an example of isolating two critical fixations (F_1 and F_x) during RA-E from a scan path on the first page and obtaining the onsets using *visualisations*.

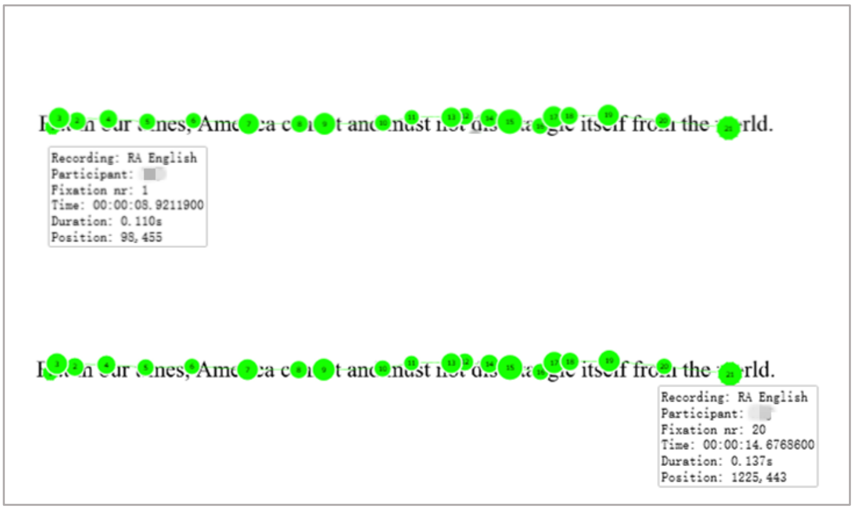


Recording: RA English
Participant: [redacted]
Fixation nr: 1
Time: 00:00:11.3069110
Duration: 0.150s
Position: 742, 686

Recording: RA English
Participant: [redacted]
Fixation nr: 6
Time: 00:00:12.4434210
Duration: 0.173s
Position: 1156, 680

Figure 5.2d: Recording the Onsets of Critical Fixations (a)

The onset of F1 is 00:00:11.307 and the onset of F_x is 00:00:12.443 according to the visualisation function. Identifying fixation onsets on page 1 is relatively straightforward, but identifying fixation onsets on page 2-4 requires taking an extra step. Figure 5.2e is an example of isolating two critical fixations (F_1 and F_x) during RA-E from a scan path on the second page and obtaining the onsets with *visualisations*.



Recording: RA English
Participant: [redacted]
Fixation nr: 1
Time: 00:00:08.9211900
Duration: 0.110s
Position: 98, 455

Recording: RA English
Participant: [redacted]
Fixation nr: 20
Time: 00:00:14.6768600
Duration: 0.137s
Position: 1225, 443

Figure 5.2e: Recording the Onsets of Critical Fixations (b)

The onset of F1 is 00:00:08.921 and the onset of F_x is 00:00:14.677 according to the visualisation function. However, since the sample sentence is on the second page of the text, 30 seconds has already past when the text on the first page was presented on the screen. Hence, the real onset of F1 is 00:00:38.921 and the onset of F_x is 00:00:44.677. In the same vein, 60 seconds should be added for fixations on page 3 and 90 seconds should be added for fixations on page 4.

To measure initial temporal EVS at S-I, the time when one subject cast his/her first fixation onto the first reading/translation unit in the sentence and the time when he/she starts either reading or sight translating the reading/translation unit are noted by the onset of the first fixation and by the articulation on-set of the first syllable. The initial temporal EVS at the sentence initials can be calculated by subtracting the former from the latter. The terminal temporal EVS at the S-T is the time span between when one subject cast his/her first fixation onto the last reading/translation unit of the sentence and when this subject starts either reading or sight translating the meaning unit. The terminal temporal EVS at S-T is calculated by subtracting the onset of F_x from the articulation onset. The onsets of the initial phonemes were marked along the timeline interface of each audio file by the Audacity software 2.2.2. An example can be seen in Figure 5.2f.

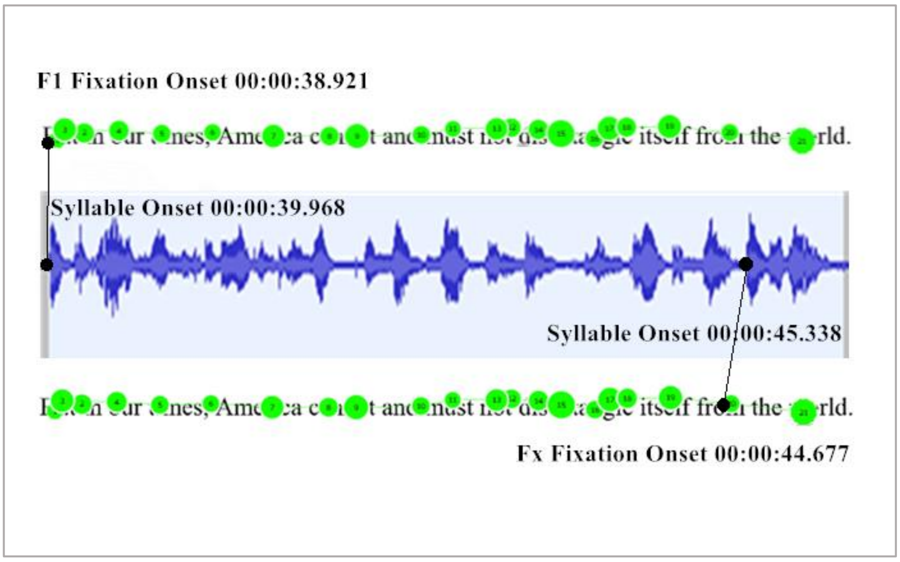


Figure 5.2f: Calculating Temporal EVS at S-I and S-T

According to (Holmqvist et al. 2011: 444), “latencies in milliseconds are the default unit with pictorial stimuli for eye-voice and voice-eye studies”. Since “the majority of latency measures operate with absolute time in milliseconds” (Holmqvist et al. 2011: 429), this study also calculates temporal EVS in milliseconds. In this example, onset-to-onset temporal EVS is 1047 ms (00:00:39.968-00:00:38.921) at S-I and 661 ms (00:00:45.338-00:00:44.677) at S-T.

There are ten sentences in the reading material of RA-E and the ST of STR E-C/E-G. Meanwhile, there are 12 sentences in the reading material of RA-C and the ST of STR C-E. In theory, for each subject, there are ten temporal EVS results at the sentence initials and ten temporal EVS at the sentence terminals in the tasks RA-E, STR E-C, and STR E-G; there are 12 temporal EVS results at the sentence initials and 12 temporal EVS at the sentence terminals in the tasks RA-C and STR C-E. However, there were a few subjects who did not succeed in starting to reading/translating the last unit on a screen page before the next screen page appeared. Moreover, subjects might also have resorted to omission when they either did not understand the meaning of the last phrase in a sentence or thought the translation of that sentence was already adequate. Under these two circumstances, it was impossible to calculate temporal EVS because there was no V (voice) to measure. However, these circumstances were infrequent.

To be specific, all subjects in Group B finished reading every sentence aloud in time. 100% of the EVS data points were registered. There are, in total, 700 EVS data points, each consists of one fixation data point and one audio data point. All subjects in Group C finished reading every sentence aloud in time. 100% of the EVS data points were registered. However, 13 data points (3.25%) were absent in STR E-G. In total, there are 400 EVS data points in RA-E and 387 EVS data points in STR E-G, each consists of one fixation data point and one audio data point. Only one subject in Group A did not manage to utter the last few syllables in RA-E, but this did not influence terminal temporal EVS calculation because the onset of F_X and the onset of articulation of the last phrase had already took place. The rest of Group A finished reading every sentence aloud both RA tasks. 100% of the EVS data points in both RA tasks were registered. There are 600 EVS data points in RA-E and 720 EVS data points in RA-C, each consists of one fixation data point and one audio data point. 20 EVS data points (3.33%) were absent in STR E-C, and 27 EVS data points (3.75%) were absent in STR C-E. Nevertheless, there are 580 EVS data points in STR E-C and 693 EVS data points in STR C-E, each consists of one fixation data point and one audio data point. In summary, 2,760 EVS results were calculated using 2,760 fixation data points and 2,760 audio data points that were registered successfully in the experiments. Despite the small amount of missing data points, each subject has two average temporal EVS results in a task, one measured by F_1 at S-Is and the other one measured by F_x at S-Ts.

In most of the cases, eye movements are moving to the right along the line when one sentence ends and another sentence follows. O’Regan (1979) found that the readers were less likely to fixate on the word *the*, compared to verbs. In fact, Carpenter and Just (1983) found that only 38% of the function words were fixated on. In the present study, it is true that when the first word at a sentence initial is *the*, it was more likely to be skipped. Nevertheless, even if the reader skips the word *the* and fixates on the following word, the EVS calculation is not affected because the present study uses sentences as the unit for calculation.

McConkie (1983: 79) showed a correlation of 0.97 between “the length of the regressions that commonly follow return sweeps of the eyes” and “the position of the immediately prior fixation relative to the left edge of the text”. This result proved that the regressive eye movements after the return sweeps did not occur randomly and researchers should not ignore them when they calculate EVS. A readers’ first fixation after a return sweep is sometimes not at the beginning of the next sentence. In a regression, the reader frequently makes a corrective movement and brings the eyes to the beginning of the sentence. This type of regressive eye movement is caused by inaccuracies in eye positioning (McConkie 1983: 79). Hence, the onset of a *wrong* fixation should not be used to measure the time the subject starts reading the sentence. Instead, the time when a subject begins to read a new sentence is noted based on the fixation cast after the corrective eye positioning.

Although many believed that “information is being extracted during the initial portion of the fixation” (Wolverton and Zola 1983: 48), some researchers have measured the critical point when the visual information from a new fixation becomes available to the visual centres of the brain (e.g., Clifton 1983; McConkie 1983; Wolverton and Zola 1983). The time it takes for the information to arrive at the brain, where the words are identified, is called the eye-mind lag (Rayner and Pollatsek 1989: 165). Having an eye-mind lag does not contradict Just and Carpenter’s (1980) immediacy assumption for three reasons. Firstly, the time lag is considerably short. Secondly, the subsequent processes take place very quickly as the “language aspects of the text must begin having their influence on processing within about 100 ms after the onset of the fixation” (McConkie 1983: 69). Thirdly, Rayner and Pollatsek (1989: 165-166) pointed out that although the visual information processing lags slightly behind a fixation, the measured time is not significantly longer than the amount of time the mind spends on processing the information because such a pattern also prolongs the availability of the information beyond the end of the fixation.

The estimation and calculation of such a time lag varied in previous studies. For example, Rayner and Pollatsek (1989: 174) argued that the eye-mind lag is as transient as 50 ms. McConkie (1983: 69) estimated the time to be “60 ms after the onset of the fixation”. Clifton (1983: 269) stated that “a reader’s eye has a period averaging 50-100 ms during which it can pick up information from text and decide to move on, and, perhaps, decide where to move”. If the so-called eye-mind lag does exist, the calculation of the EVS in this study is slightly inflated. It could make a difference, although very small, in the measurement of the temporal EVS because the calculation of EVS is based on the exact time when fixations start. Nevertheless, when calculating the length of the temporal EVS during RA and STR, the present study did not subtract the alleged 50-100 ms from the fixations for three reasons. Firstly, there was not a significant amount of literature that supported this measurement. Secondly, even if there had been, 50-100 ms is a very short time span that does not alter the results drastically. Thirdly, if the measured EVS length is generally inflated by 50-100 ms, it does not alter the results when EVS is compared within or across groups.

5.3 Eye-tracking Data Acquisition

5.3.1 Eye-tracking Data Quality

Eye-tracking data quality is a property of the data samples collected by the eye-tracker. In any normal data set, some values fall outside the range and they are referred to as noise. These outliers are potentially deviant subjects/items/data that should be identified. Holmqvist *et al.* (2011: 182) summarised the categories of eye-tracking data noise as following: artefacts, data loss, and blinks. Since "poor data quality can never be remedied by later data processing and statistical analysis" (Holmqvist *et al.* 2011: 144), the process of removing any variations that did not derive from true eye movements from the data set should be completed before any analysis takes place. The proportion of data discarded due to poor data quality varies in different studies. The discard rate was suggested to be 2%-5% by Holmqvist *et al.* (2011), whereas some other researchers, such as Schnipke and Todd (2000) and Pernice and Nielsen (2009) reported a higher percentage of data loss over 20%.

According to Holmqvist *et al.* (2011: 29), McConkie (1981) argued that "every published research article should list measured values for the quality of the data used" but this has not yet become the standard after three decades (Holmqvist *et al.* 2011: 29). In the field of translation studies, Jensen (2008) was concerned about the overall accuracy of the eye-tracking data collected by the eye-tracking equipment a decade ago. Sadly, because of the complicated process and the lack of guidance, the eye-tracking data quality assessment is often neglected even though it is a crucial step before data analysis (Hvelplund 2014: 216).

In some cases, researchers did eliminate some eye-tracking recordings but did not justify their assessment. For instance, Sharmin *et al.* (2008: 36) discarded the data collected from three out of 21 subjects due to "poor eye-tracking quality", but they did not specify their quality assessment process or criteria. Three out of eight subjects in O'Brien's (2008: 82) study were removed because their eye-movements were "not adequately tracked", but O'Brien did not specify her measurements. In some other cases, the method used to judge the data quality was only one-fold. For example, Jensen (2008) assessed the quality of the eye-tracking data with the ClearView gaze replay function. According to him, the researchers could have a rough idea of the percentage of artefact eye movements and judge the quality of the data by observing and noting down significant traces of the *Brownian motions*, which are rapid vertical or horizontal eye movements (Jensen 2008: 168). This method, however, has not been cross-checked with any other criterion.

Holmqvist *et al*. (2011), considered this issue as it was still waiting to be addressed. They even argued that having the skill and knowledge to record high-quality eye-tracking data should be learned through training (2011: 144). As a result, one of the many challenges for researchers who wish to carry out translation process studies is that they "have to use their own judgments and should take care to report the thresholds used to filter out low-quality data" due to a lack of set standards and terminologies (Saldanha and O'Brien 2014: 143).

In this research, I believe that a systematic eye-tracking data quality assessment is of great importance. Therefore, the present research aims to set its own eye-tracking data quality assessment criteria based on the research design and the experiment conditions. The following sections explore the criteria employed by other studies; establish the set of criteria used by the current eye-tracking research; present the results of the eye-tracking data quality assessment; and then briefly discuss some results from the assessment.

5.3.2 Quality Assessment of Eye-tracking Data: A Brief Review

Assessing the quality of eye-tracking data and then eliminating those data recordings that are low quality is the first step. Researchers using the Tobii Studio to analyse the eye-tracking experiments can easily find a quality rating called the *sample* in the *replay* window. Such a quality rating comes in percentages. As Saldanha and O'Brien (2014: 142) suggested: if the percentage is 100%, it means that "both eyes were found consistently throughout the recording"; if it is 50%, it "means that either one eye was found for the full recording or both eyes during half the time". This function is particularly useful to discover *binocular disparity*, which means the subject has a dominant eye and a passive eye (Holmqvist *et al*. 2011: 24). Some researchers have used this sample to judge the adequacy of data collection. For instance, Walker (2018) set the passing sample rates at 75%. On the other hand, some researchers have used their own eye-tracking data quality assessment criteria. A couple of eye-tracking based translation studies and their eye-tracking data quality assessment are described and evaluated here in chronological order.

Based on Rayner's (1998: 373) finding that the mean fixation duration

(MFD) during silent reading is usually approximately or more than 225 ms, Pavlović and Jensen (2009) discarded eye-tracking recordings that contained a high percentage of fixations shorter than 200 ms. In the same year, Jensen, Sjørup, and Balling (2009) also used this standard for quality assessment of eye-tracking data. They took an average fixation below 175 ms to indicate a measurement error and discarded data from the subjects with a MFD shorter than 175 ms. Jensen, Sjørup, and Balling (2009), however, encountered a discard percentage as high as 62.5% based only on this data quality parameter. They reflected on this extraordinarily high discard rate and regarded it as a technical problem caused by setting the viewing distance from the monitor between 55 cm to 80 cm. Thus, they strongly recommended that future researchers should set the viewing distance as 55 cm to 60 cm (Jensen, Sjørup, and Balling 2009: 326).

Hvelplund (2011) also assessed his eye-tracking data collected in a written translation experiment with the help of the MFD. Hvelplund's (2011) MFD threshold was set at 200 ms according to Pavlović and Jensen's (2009) report on an eye-tracking experiment, which had discarded 50% of the eye-tracking recordings based on this criterion. Hvelplund (2011: 106) argued that "recordings in which the mean fixation duration (MFD) was lower than 200 ms could be (partially) corrupted in that the short fixations could reflect noisy data rather than a low amount of cognitive effort invested during a given task". Since Hvelplund (2014: 217) argued that MFD "is a relatively crude measure, which ignores the potential difference in *completeness* of eye-tracking recordings", he introduced two extra criteria to assess the eye-tracking data quality: the Gaze Time on Screen (GTS) and Gaze Sample to Fixation percentage (GSF/GFP).

The GTS percentage indicates the subjects' gaze time (total fixation duration) on the whole screen as a percentage of the total production time. It is dedicated to examining the *completeness* of eye-tracking recordings, which symbolises "how much eye movement has been successfully recorded by the eye tracker compared to how much has not been recorded" (Hvelplund 2014: 217). If the participant spent too little time looking at the screen, the GTS percentage would be very low and there would not be a sufficient amount of data to be analysed. The mean GTS in Hvelplund's (2011) study was 55.7%, because subjects spent a significant amount of the task time looking at the keyboard.

Then, Hvelplund (2011) used the Gaze Sample to Fixation percentage (GSF/GFP) to gauge eye-tracking data quality. A threshold of 75% was set, as the theory behind this measurement was that "Comparing typical saccade duration with fixation duration, saccades account for about 5-15 percent of all eye movements in reading, while the remainder 85-95 percent are fixations" Hvelplund (2011: 67). The GSF/GFP score was calculated with the formula: [(number of gaze samples ÷ number of fixation gaze samples) × 100]. According to Hvelplund (2014: 218), "in a recording containing 11,000 gaze samples, of which 10,000 belong to fixations, the GSF percentage is 90.9". Unfortunately, there seems to be a mismatch between the measurement and the theory behind it. It was clearly the "duration" of the fixations and saccades that matters in theory, but Hvelplund (2011/2014) has calculated the GSF/GFP with the "number" of the fixations and saccades. Ideally, if the event-detection algorithm is sufficient, the number of the saccades would be equal to or plus/minus 1 the number of the fixations on a still stimulus. When the stimulus is not static, the number of the fixations would be lower (Holmqvist *et al*. 2011: 403). This explains why Hvelplund (2011) has set the threshold at 75%, regardless of the principle that if the percentage was lower than 85%, the data should have been outliers of the normal eye-tracking data. In summary, the subjects' data was discarded when it failed to meet two or three of the quality assessment criteria (MFD, GTS, GSF/GFP). In the end, the discard percentage was 11.1% in Hvelplund's (2011) research.

Thereafter, Sjørup (2013) applied two of Hvelplund's (2011) criteria: GTS and MFD. While Hvelplund (2011) set the GTS threshold at around 30%, Sjørup (2013) raised the standard to 50% based on an average GTS score of 65% achieved in her research. On the other hand, Sjørup (2013) lowered the threshold of MFD to 180 ms. The eye movements from all subjects were collected and tested against these two standards. Since all the eye-tracking recordings passed the quality assessment, no recording was excluded from further analysis (Sjørup 2013: 106).

Schmaltz *et al*. (2016) used three data quality assurance criteria to ensure consistency in the samples. Following Hvelplund (2011) and Sjørup (2013), Schmaltz *et al*. (2016) established the threshold of the MFD at a minimum of 180 ms. The GTS threshold was set at 30%. Unfortunately, Schmaltz *et al*. (2016) did not specify the reason or the foundation for setting this threshold. Schmaltz *et al*. (2016) designed a third criterion called *% of valid win gaze data*. Specifically, if one fixation was on the ST, win = 1; if one fixation was on the TT, win = 2; if it was on neither ST nor TT, win = 0. The *% of valid win gaze data* was thus calculated by dividing the number of occurrences of win = 1 plus win = 2 by the total number of wins. They "arbitrarily established" the threshold at ≥ 40% (Schmaltz *et al*. 2016: 244). This criterion, fundamentally, valued the number of fixations on the TT more than that on the ST. However, no reason was given. Nevertheless, this criterion does not apply to STR studies because the TT during STR comes in an audio form.

Based on the relevant literature, Zheng and Zhou (2018) adopted MFD as one of the eye-tracking data quality assessment criteria. The threshold was set at 200 ms. The 24 subjects' MFD results were calculated and only one out of the 24 subjects had a MFD result lower than 200ms. The FTS (Fixation Time on the ST as a percentage of the total task time) was inspired by Hvelplund's (2011) GTS and created in comparison with the *sample* score provided by the Tobii Studio. The threshold was set at one standard deviation below the group's average FTS and four recordings did not meet this criterion. Time of the Unclassified Sample as a Percentage of total Task time (USP), which reflects the percentage of the unusable eye-movement data, was established as the other eye-tracking data quality assessment criterion. The amount of the unclassified data sample was measured and exported with the help of the I-VT filter. A USP rate higher than one standard deviation above the group's average USP was considered an outlier outside the normal sample distribution. Three eye-tracking recordings failed to meet this requirement. In the end, the discard percentage was 20.83%.

Another type of data quality management operates on a micro level. Instead of carrying out a data quality assessment and then excluding the recordings deemed unfit, it merely deletes all the outliers in each recording. For instance, Schaeffer *et al.* (2016) excluded the data-points which were more than 2.5 standard deviations above or below the average of a dependent variable. This resulted in an exclusion of less than 5% for all the dependent variables in their study. Compared to the aforementioned data quality assessment, this type of data quality management has its merits: 1) by deleting the marginal data-points in each eye-tracking log, the number of eye-tracking logs and subjects are kept intact; 2) the remaining data-points are all qualified. However, it also has its drawbacks: 1) the marginal data-points are not necessarily errors. Sometimes, they could provide valuable insight into some specific issues; 2) since the means and standard deviations of a particular dependent in each recording vary, a different percentage of data of each subject is kept for analysis, thus bringing in another variable; 3) even if a subject's recording is of very poor quality, it is still included in the analysis with a very low percentage of data-points.

5.3.3 Four Quality Assessment Criteria

Based on the pilot study, the present study applied four quality assessment criteria: **F**ixation **T**ime on **S**ource text as a percentage of total task time (FTS), Time of the **U**nclassified **S**ample as a **P**ercentage of total Task time (USP), **M**ean **F**ixation **D**uration (MFD), and **S**accade Duration as a **P**ercentage of Pure **G**aze Activities (SPG). Although the researcher could not troubleshoot all accuracy and precision problems from the beginning of the experiment, these four measurements operate and coordinate with each other to address different types of inaccurate eye-tracking data classified by Jensen (2008: 168): undetected fixations, incorrectly detected fixations, and variations of these two basic types.

Researchers working with statistical computing, such as Crawley (2002), argued that outliers that are greater than 2.5 standard deviations from the mean should be removed. Subsequently, researchers interested in translation studies (e.g., Macizo and Bajo 2004, Jensen, Sjørup, and Balling 2009) followed this advice and filtered out data points that are greater than 2.5 standard deviations from the average result. Of course, there were researchers who chose to raise or lower the bar by manipulating how many standard deviations to add to or subtract from the average results. For example, Tabachnick and Fidell (2000) excluded the values that were more than 3.29 standard deviations above or below the mean in their study. In the eye-tracking study conducted by Krieber *et al*. (2017), one subject was eliminated based on the threshold set at three standard deviations above the sample's mean MFD. Zheng and Zhou (2018) set the criterion threshold at one standard deviation below the mean FTS and one standard deviation above the mean USP. Based on the literature and the most recognised normal data distribution used in statistical computing, the present study applied mean ± 2.5 standard deviations for FTS and USP.

5.3.3.1 Fixation Time on Source Text as a Percentage of Total Task Time

In contrast with other written translation experiments, the reading task in the present research took 120 seconds or less and the STR task took 200 seconds or less, during which the subjects' eyes focused mainly on the screen instead of the keyboard. Although there were four pages of reading material and ST, they were displayed one after another, all situated in the centre of the screen. As a result, occasional fixations away from the valid area were considered noise in the experiments. Fixation Time on Source text as a percentage of total task time (FTS) was set as the first criterion because it indicates the amount of valid fixation time during the task. FTS = [(total fixation duration on the ST ÷ total task time) × 100].

The term "total gaze time" in the relevant literature (e.g., Jakobsen and Jensen 2008; Hvelplund 2011; Hvelplund 2014) refers to the "total duration of all the fixations during execution of the task" (Jakobsen and Jensen 2008: 107). The "Fixation Time" here is the same as the term "Gaze Time" (Hvelplund 2011), indicating the total fixation duration. Total fixation duration increases as fixation count and MFD increase. The ST area was drawn to be the *Area of Interest* (AOI) and only the fixations that fell in this AOI were considered valid. Regarding TTT, it was different for every individual. Some of them completed reading or sight translating before the set time had passed, whereas some others failed to accomplish the job in the allocated time. A subject's TTT consisted of the time spent on all four pages, each from the moment he/she first fixated on the first phrase on a page to the moment he/she finished reading/translating the content on that page.

A low FTS score indicates that the subject looked away from the ST on the screen for a considerable amount of time and/or the eye tracker failed to capture most of the eye movements within the AOI. Used by Hvelplund

(2011), the GTS (Gaze Time on the Screen as a percentage of total production time) indicated the subjects' total fixation duration on the whole screen as a percentage of total production time. The two criteria seem similar, but the difference between the GTS and the FTS is that the former includes fixation on the entire screen while the latter focuses only on the specific eye-fixations on the ST. The GTS was calculated in the pilot study to check the difference between the eye-tracking data obtained during written translation and STR. It turned out that for the 19 subjects whose data had passed the data quality assessment in the pilot study, the mean GTS was 80.36%. This GTS score was much higher than the result of Hvelplund's (2011) written translation study (55.70%). It was mainly due to the format of the language output during STR. The high GTS in the pilot study, therefore, indicated that the eye tracker could provide a large amount of information for studies on reading-speaking modalities including the RA and the STR. Furthermore, it was not considered appropriate to use Hvelplund's (2011) relatively low GTS threshold in RA and STR experiments.

In summary, the FTS threshold was set at 2.5 standard deviations below the average score for individual tasks. That means every test group had its own FTS passing scores in each task. For instance, Group A's FTS pass threshold in STR E-C did not have any impact on Group C's FST pass threshold in RA-E.

5.3.3.2 Time of the Unclassified Sample as a Percentage of Total Task Time

The present research then assessed the data quality based on the percentage of the non-fixation and non-saccade data samples. The "time of the Unclassified Sample as a Percentage of total task time (USP)" was set as the second criterion. USP = [(time of the unclassified eye movement sample ÷ total task time) × 100].

The I-VT Filter exported data classified as fixations, saccades, and the unclassified eye movements. The unclassified samples were regarded as noise, referring to those eye movements that could not be classified as either fixations or saccades. They were usually generated when the subjects' eyes were closed or blocked by something. Hence, the higher the USP, the lower the quality of the eye-tracking data. When Hvelplund (2014: 218) applied

GTS in eye-tracking data quality assessment, he pointed out that a low GTS was not necessarily a result of poor data quality because it could also be due to the substantial periods of time during which the subject looked away from the monitor. However, applying the USP along with the FTS could solve this problem.

In summary, as mentioned above, the USP threshold was set to be at 2.5 standard deviations above the average score for individual tasks. Every test group had its own USP passing scores in each task. For example, Group A's USP pass threshold in RA-E did not have an impact on their USP pass threshold in STR E-C.

5.3.3.3 Mean Fixation Duration

The third criterion was Mean Fixation Duration (MFD). MFD = (total fixation duration on the ST ÷ total fixation count). Rayner (1998: 373) pointed out that the mean fixation duration is 225 ms during silent reading and proposed a threshold for MFD: a recording with a mean fixation duration shorter than 200 ms should be considered invalid for further analysis (ibid.: 106). In other studies, different MFD thresholds were adopted, ranging from 180 ms (Sjørup 2013: 105) to 200 ms (e.g., Pavlović and Jensen 2009; Hvelplund 2014).

However, in contrast with Hvelplund's (2011) MFD calculated by dividing the combined fixation duration on the screen by the number of fixations on screen, the MFD in the present study was calculated by dividing the combined fixation duration on the ST by the number of fixations the ST.

This research adopted 200 ms as the MFD threshold, which is in line with Pavlović and Jensen (2009) and Hvelplund (2014).

5.3.3.4 Saccade Duration as a Percentage of Pure Gaze Activities

The last criterion is **S**accade Duration as a **P**ercentage of Pure **G**aze Activities (SPG). The concept *Gaze Samples* includes fixations, saccades, and samples that are not accounted for (Hvelplund 2011). The concept *Pure Gaze Activities*, on the other hand, refers to the sum of pure fixations and saccades. In other words, pure gaze activities are gaze samples excluding the noise. SPG = [(saccade duration ÷ the sum of fixation and saccade duration) × 100].

Many researchers have found evidence for the hypothesis that the length of fixations and saccades are related to the processing time of the reading task (e.g., Shebilske 1975). For instance, based on the finding that saccades generally last between 20 and 35 ms (Rayner and Pollatsek 1989: 113), Jakobsen and Jensen (2008: 113) concluded that saccades "consequently constitute some 10 to 15% of reading time". Since the reading time consists of the time spent on blinking, closing one's eyes, and looking away from the text, the present study proposes a more precise measure of the proportion of the fixations and saccades to establish an alternative criterion.

Although the length of the saccades and duration of the fixations might vary in accordance with the readers' reading skills (Downing and Leong 1982: 144-5), the proportions of the fixations and the saccades were said to fluctuate at around 94% and 6% respectively by Gibson and Levin (1975: 354). As mentioned above, Hvelplund (2011) established the criterion Gaze Sample to Fixation percentage (GSF/GFP) on the foundation of the fixation-saccade proportions, which are 5-15% and 85-95%. Although his calculation method turned out to be wrong, it certainly inspired the invention of the last eye-tracking data quality assessment criterion in the present study.

This fixation-saccade ratio in STR was not used as a quality assessment criterion in the pilot study because researchers (Jakobsen and Jensen 2008) had acknowledged that the reading patterns during *reading for comprehension* and *reading for translation* were different. Also, the eye movement patterns and characteristics during STR might be influenced by the simultaneous oral production process. Instead, saccade duration as a percentage of the total pure gaze activities during STR was tested by the pilot study, in order to see whether this fixation-saccade ratio was also within the range in the STR task. In the pilot study, results from the qualifying 19 subjects showed that the mean SPG was 9.96%, with a standard deviation of 2.08%. The range of SPG mean ± 2.5 standard deviations was between 4.76% and 15.16%. The average result was roughly between 5% and 15%. Not a single participant's pilot results were beyond the range.

Thus, it was proposed that the saccades consequently constitute approximately 5% to 15% of the duration of all of the eye movements. Hence, this criterion was applied in the present study for both the RA and the STR tasks in the eye-tracking experiment.

5.3.4 Quality Assessment of Eye-tracking Data in this Research

Although "the eye tracker manufacturers claim that contact lenses and glasses for normal corrected vision do not have an impact on the collection of gaze data" (Saldanha and O'Brien 2014: 139), the researcher of the present study found that the eye-tracking data elicited from subjects who were wearing thick glasses tended to be poor. Two of the recruited subjects whose glasses were very thick failed to complete the calibration and were eliminated from the experiment. Saldanha and O'Brien (2014: 139) acknowledged that "even when calibration is fine, subsequent data capture can be poor". Therefore, researchers need to carry out a data quality assessment before analysing the data, rather than assuming that the data is of a high quality as long as the subject has passed the calibration. To date, no standard assessment criterion has been developed and researchers tend to set up their own measurements and rules. The criteria for eye-tracking data quality assessment in the present research experiment were explained in the previous section. They are:

- FTS: Fixation Time on Source text as a percentage of total task time [(total fixation duration ÷ total task time) ×100]. The threshold is 2.5 standard deviations below the mean FTS score of each task;
- USP: The time of the Unclassified Sample as a Percentage of total task time [(the unclassified data sample duration ÷ total task time) ×100]. The threshold is 2.5 standard deviations above the mean USP score of each task;
- MFD: Mean Fixation Duration [total fixation duration ÷ fixation count]. The threshold is 200 ms;
- SPG: The standard range of the Saccade duration as a Percentage of pure Gaze activities [(total saccade duration ÷ pure gaze activities duration) ×100]. The standard is between 5% and 15%.

The present research strives to standardise the parameters for the inclusion and exclusion of the eye-tracking data by setting the standards and then discarding the recordings of poor data quality based on these standards. The general statistic threshold (mean ± 2.5 standard deviations) was used for FTS and USP, following previous researchers such as Macizo and Bajo (2004), Jensen, Sjørup, and Balling (2009). The thresholds were set based on these studies for MFD and SPG.

The data needed to comply with at least three of the four eye-tracking data assessment criteria to ensure that the analysis and its results were not skewed by flawed data. If an eye-tracking record failed to meet more than one of the assessment criteria above, the recording was not further analysed. Furthermore, if any of a participant's recordings failed to pass the data quality assessment, he/she was not included as a formal subject, regardless of the quality of his/her other recordings. The results of the eye-tracking data quality assessment were:

Group A

Except for two subjects who failed to pass the calibration, 31 Chinese subjects took part in this research experiment. As mentioned above, the MFD threshold is 200 ms and the SPG standard range is between 5% and 15%. The FTS and the USP thresholds, which relied entirely on the mean score and the standard deviation in each task are shown in Table 5.3a.

Table 5.3a: FTS and USP Thresholds of Group A in RA and STR

Group A	RA-C		RA-E		STR C-E		STR E-C	
	FTS	USP	FTS	USP	FTS	USP	FTS	USP
μ	0.820434	0.054771	0.825060	0.048949	0.748666	0.073126	0.758769	0.069915
SD	0.046850	0.032492	0.040769	0.029048	0.069656	0.037169	0.061379	0.036490
$\mu \pm 2.5SD$	0.703310	0.136000	0.723137	0.121568	0.574526	0.166048	0.605322	0.161141

One subject's eye-tracking recording did not pass the data quality assessment. Both his/her STR tasks failed to meet the following three criteria: MFD (STR E-C: 0.1489; STR C-E: 0.1429), FTS (STR E-C: 0.5579; STR C-E: 0.4605), and SPG (STR E-C: 0.1961; STR C-E: 0.229). Therefore, this participant was excluded. As a result, Group A consisted of 30 Chinese subjects whose data passed the data quality assessment. The quality assessment results, including FTS, USP, MFD, and SPG in all of the tasks, are presented in Table 5.3b.

Table 5.3b: Data Quality Assessment of Group A in RA and STR

Group A	RA-E				STR E-C				RA-C				STR C-E			
	FTS	USP	MFD	SPG	FTS	USP	MFD	SPG	FTS	USP	MFD	SPG	FTS	USP	MFD	SPG
A01	0.837699	0.031979	0.265804	0.086227	0.815415	0.029827	0.244694	0.108842	0.867299	0.032768	0.314433	0.072666	0.786778	0.041980	0.254622	0.118663
A02	0.810060	0.048353	0.229535	0.098634	0.790784	0.039432	0.245963	0.128710	0.810046	0.050517	0.255867	0.097899	0.808721	0.044223	0.282607	0.110840
A03	0.899496	0.000246	0.291673	0.087235	0.783157	0.088908	0.267832	0.115686	0.893020	0.006881	0.280194	0.087342	0.863382	0.034688	0.310592	0.089278
A04	0.860702	0.028474	0.256103	0.094727	0.811644	0.034474	0.253352	0.130834	0.843852	0.033652	0.253352	0.130834	0.801199	0.028413	0.261625	0.103423
A05	0.800148	0.122890	0.241118	0.083796	0.632060	0.192914	0.226549	0.115948	0.781453	0.112983	0.230192	0.088296	0.673765	0.134759	0.234438	0.111965
A06	0.838018	0.027980	0.252768	0.089746	0.749020	0.043483	0.207198	0.134118	0.814566	0.033871	0.254196	0.102844	0.788708	0.037021	0.242876	0.106893
A07	0.841761	0.036433	0.243725	0.093988	0.728466	0.076695	0.208678	0.144252	0.831964	0.051268	0.263870	0.087985	0.774715	0.047069	0.246704	0.120494
A08	0.840187	0.054705	0.251131	0.091232	0.724845	0.100661	0.223755	0.135491	0.850046	0.039347	0.261756	0.086855	0.745551	0.061541	0.242872	0.124298
A09	0.900477	0.010813	0.291129	0.067417	0.822964	0.035907	0.250308	0.110716	0.902739	0.015434	0.320248	0.058349	0.761163	0.057856	0.269485	0.112012
A10	0.820221	0.057206	0.255709	0.109636	0.757061	0.110805	0.332022	0.117424	0.737374	0.117852	0.237074	0.120923	0.725592	0.141528	0.340846	0.103453
A11	0.834425	0.040332	0.242985	0.103005	0.815233	0.056416	0.300264	0.112055	0.750524	0.126513	0.244538	0.123744	0.825522	0.064246	0.374448	0.093321
A12	0.812358	0.052885	0.243265	0.102466	0.805802	0.043246	0.245230	0.119197	0.850905	0.031462	0.263151	0.091441	0.804624	0.044322	0.240972	0.120493
A13	0.847342	0.030915	0.250083	0.091971	0.634075	0.092194	0.199357	0.139124	0.821086	0.057771	0.258000	0.080279	0.607534	0.110417	0.203057	0.124361
A14	0.872665	0.016640	0.262794	0.086525	0.881376	0.023797	0.375842	0.077558	0.846841	0.033745	0.268043	0.084174	0.854809	0.046477	0.384336	0.073310
A15	0.853640	0.028114	0.242367	0.089725	0.835642	0.030702	0.243698	0.129983	0.862111	0.032014	0.268774	0.068198	0.816120	0.031072	0.263509	0.127188
A16	0.733115	0.107926	0.230873	0.111567	0.701900	0.114473	0.249585	0.111495	0.769141	0.121424	0.258594	0.078136	0.727789	0.107774	0.251607	0.136513
A17	0.733292	0.123751	0.235380	0.108086	0.708994	0.094049	0.241946	0.129988	0.765133	0.008365	0.250072	0.112666	0.647561	0.159558	0.247296	0.134649
A18	0.764417	0.040293	0.221246	0.119247	0.788408	0.030895	0.222417	0.137082	0.859232	0.030166	0.284008	0.083875	0.790008	0.034570	0.260433	0.116936
A19	0.823174	0.047461	0.242656	0.087430	0.771168	0.060606	0.226798	0.114762	0.830761	0.055414	0.268082	0.087704	0.759991	0.057692	0.252852	0.106080
A20	0.829963	0.057874	0.260926	0.081758	0.731624	0.096325	0.250679	0.116044	0.827201	0.058693	0.270552	0.075672	0.736415	0.099639	0.263077	0.111379
A21	0.837600	0.039714	0.240536	0.097876	0.669466	0.078442	0.202807	0.171633	0.729687	0.080579	0.213221	0.127387	0.644102	0.082019	0.186922	0.142929
A22	0.822797	0.049619	0.272811	0.075858	0.763084	0.068463	0.272542	0.089173	0.841133	0.054388	0.318479	0.062293	0.804108	0.043644	0.309582	0.078722
A23	0.815813	0.046831	0.260141	0.102755	0.695728	0.096019	0.202323	0.142346	0.753787	0.085493	0.228587	0.107559	0.662704	0.079687	0.199769	0.146456
A24	0.802179	0.070464	0.240730	0.093875	0.741607	0.071989	0.206696	0.126283	0.808918	0.050431	0.237441	0.097345	0.690934	0.091556	0.217277	0.129938
A25	0.762380	0.055588	0.227778	0.107898	0.735948	0.042912	0.212586	0.127145	0.770597	0.061063	0.216955	0.107618	0.731057	0.054548	0.245395	0.100570
A26	0.848862	0.040363	0.245240	0.090238	0.832521	0.034188	0.239825	0.113971	0.863230	0.050172	0.268185	0.072429	0.616487	0.146345	0.219663	0.212356
A27	0.804751	0.085140	0.256063	0.103022	0.741865	0.102016	0.257889	0.126106	0.812922	0.083717	0.246402	0.107432	0.734414	0.102412	0.315566	0.093398
A28	0.876413	0.023956	0.273741	0.081829	0.830783	0.042286	0.247919	0.108558	0.895779	0.017875	0.287987	0.072413	0.823318	0.060587	0.267969	0.104860
A29	0.836561	0.042496	0.257126	0.079833	0.779690	0.083102	0.267437	0.098666	0.821450	0.056425	0.272610	0.082272	0.773604	0.077533	0.293932	0.090088
A30	0.791288	0.049026	0.250112	0.086732	0.682729	0.082228	0.212057	0.120343	0.800220	0.052835	0.265766	0.069925	0.679311	0.070601	0.211170	0.114661
μ	0.825060	0.048949	0.251185	0.093478	0.758769	0.069915	0.244608	0.121784	0.820434	0.054771	0.262021	0.090885	0.748666	0.073126	0.263183	0.115318
SD	0.040769	0.029048	0.016727	0.011568	0.061379	0.036490	0.038756	0.017696	0.046850	0.032492	0.026154	0.019497	0.069656	0.037169	0.047375	0.025436

Group B

36 British English-speaking subjects took part in this research experiment. The MFD threshold is 200 ms and the SPG standard range is between 5% and 15%. However, since the FTS and the USP thresholds relied entirely on the mean score and the standard deviation in each task, they were not the same as Group A's FTS and USP thresholds in RA-E. The results are shown in Table 5.3c.

Table 5.3c: FTS and USP Thresholds of Group B in RA-E

Group B	RA-E	
	FTS	USP
μ	0.837468	0.033444
SD	0.040512	0.024320
$\mu \pm 2.5SD$	0.736188	0.094244

One participant's eye-tracking recording did not pass the data quality assessment. The RA-E task of his/hers failed to meet three of the criteria: FTS (0.3940), USP (0.2887), and SPG (0.2609). Therefore, this participant's recordings were excluded from further analysis. As a result, Group B consisted of 35 British subjects whose data passed the data quality assessment. The quality assessment results are presented in Table 5.3d.

Table 5.3d: Data Quality Assessment of Group B in RA-E

Group B	RA-E			
	FTS	USP	MFD	SPG
B01	0.821173	0.053290	0.208528	0.125674
B02	0.774265	0.035934	0.260286	0.162061
B03	0.772443	0.070717	0.215505	0.119917
B04	0.835141	0.020359	0.216466	0.118636
B05	0.808332	0.041537	0.201336	0.119566
B06	0.906330	0.002949	0.265735	0.072531
B07	0.771940	0.065537	0.220085	0.111047
B08	0.790780	0.080326	0.239270	0.117807
B09	0.881905	0.001395	0.252280	0.105200
B10	0.874050	0.003094	0.210000	0.102965
B11	0.815972	0.062806	0.216726	0.112982
B12	0.866100	0.062151	0.232662	0.068939
B13	0.848223	0.050374	0.235519	0.108091
B14	0.851178	0.022270	0.221713	0.077344
B15	0.864805	0.004325	0.238451	0.116215
B16	0.828073	0.052495	0.239202	0.105609
B17	0.878960	0.006648	0.271632	0.078650
B18	0.862554	0.020377	0.289067	0.081577
B19	0.758001	0.038237	0.233766	0.172346
B20	0.790497	0.064057	0.200548	0.133378
B21	0.872337	0.027837	0.308827	0.086614
B22	0.830287	0.024073	0.197107	0.118210
B23	0.848620	0.006891	0.220200	0.127074
B24	0.877200	0.016485	0.261625	0.087470
B25	0.868826	0.006550	0.215263	0.085527
B26	0.885672	0.004636	0.241296	0.092347
B27	0.890121	0.015080	0.284115	0.086673
B28	0.830783	0.033739	0.223224	0.111415
B29	0.785395	0.053366	0.238224	0.111912
B30	0.813309	0.054837	0.219269	0.106393
B31	0.866667	0.008717	0.216996	0.096985
B32	0.812797	0.057655	0.221593	0.099296
B33	0.895719	0.009083	0.262691	0.084144
B34	0.832460	0.025875	0.199254	0.117847
B35	0.800476	0.066836	0.200278	0.115448
μ	0.837468	0.033444	0.233678	0.106797
SD	0.040512	0.024320	0.027680	0.022608

Group C

21 native German-speaking subjects took part in the research experiment. Similarly to the other test groups, the MFD threshold is 200 ms and the SPG standard range is between 5% and 15%. However, the FTS and the USP thresholds relied entirely on the mean score and the standard deviation in each

task. The results are shown in Table 5.3e.

Table 5.3e: FTS and USP Thresholds of Group C in RA-E and STR E-G

Group C	RA-E		STR E-G	
	FTS	USP	FTS	USP
μ	0.828477	0.042229	0.769699	0.090237
SD	0.044890	0.021461	0.075571	0.049513
μ ± 2.5*SD*	0.716253	0.095882	0.580772	0.214020

Only one participant's eye-tracking recording did not pass the data quality assessment. Although his/her RA-E had passed the assessment, the STR E-G task of his/hers did not meet any of the four criteria: MFD (0.1733), FTS (0.4561), USP (0.2851), and SPG (0.1849). Therefore, the participant was excluded from the group. As a result, Group C consisted of 20 German students whose data had passed the data quality assessment. The quality assessment results are presented in Table 5.3f.

Table 5.3f: Data Quality Assessment of Group C in RA-E and STR E-G

Group C	RA-E				STR E-G			
	FTS	USP	MFD	SPG	FTS	USP	MFD	SPG
C01	0.847791	0.035961	0.231968	0.117784	0.819266	0.057702	0.289075	0.110462
C02	0.879838	0.010668	0.249794	0.080376	0.807497	0.067133	0.279010	0.096066
C03	0.841368	0.049549	0.270458	0.088482	0.759848	0.069226	0.258928	0.128444
C04	0.858478	0.030267	0.236397	0.087655	0.840179	0.053771	0.242432	0.090502
C05	0.852830	0.028914	0.265197	0.099100	0.790127	0.075334	0.321963	0.092517
C06	0.873911	0.029402	0.208163	0.095219	0.843853	0.024270	0.223588	0.122811
C07	0.859432	0.037099	0.260637	0.087982	0.803746	0.094194	0.292243	0.090470
C08	0.768084	0.076155	0.210574	0.135060	0.665446	0.141843	0.223977	0.128300
C09	0.790313	0.024608	0.249630	0.114331	0.819957	0.122125	0.239724	0.121323
C10	0.848669	0.039305	0.239733	0.105409	0.620803	0.216568	0.217109	0.136373
C11	0.744531	0.092911	0.191893	0.138830	0.758321	0.128495	0.227837	0.142070
C12	0.811323	0.057598	0.215464	0.084532	0.611734	0.160302	0.188705	0.139955
C13	0.835911	0.053086	0.230491	0.087069	0.704463	0.102908	0.225285	0.116820
C14	0.875346	0.023878	0.246875	0.089241	0.850520	0.033668	0.348292	0.108738
C15	0.774700	0.076605	0.198684	0.122812	0.718834	0.122007	0.202353	0.142176
C16	0.873759	0.018531	0.247934	0.066545	0.888689	0.014097	0.305683	0.082054
C17	0.860171	0.038377	0.245809	0.099921	0.804139	0.078313	0.279973	0.096731
C18	0.829947	0.026632	0.215714	0.121994	0.800172	0.058764	0.254152	0.127150
C19	0.732431	0.061007	0.189248	0.141364	0.723422	0.117585	0.222351	0.132891
C20	0.810713	0.034031	0.249556	0.135472	0.762961	0.066442	0.274010	0.131221
μ	0.828477	0.042229	0.232711	0.104959	0.769699	0.090237	0.255834	0.116854
SD	0.044890	0.021461	0.024264	0.022063	0.075571	0.049513	0.041989	0.019493

In summary, the subjects whose eye-tracking recordings failed the data quality assessment were excluded from further analysis. As a result, there were 30 subjects in Group A (A01-A30), 35 subjects in Group B (B01-B35), and 20 subjects in Group C (C01-C20). The overall discard rate was 5.55%, including the two Chinese subjects excluded in the calibration stage. The overall discard rate of the eye-tracking data quality assessment based on the four criteria was 3.41%. Also, the average Tobii Studio sample rate was 89.63%, which is considered high.

The discussions in the following chapters are based solely on the subjects whose eye-tracking recordings have passed the data quality assessment. One of the criteria in the data quality assessment, namely the USP, is not analysed further. Since it only tells us the percentage of the recorded eye-movements that were considered as noise, it does not have anything to do with the subjects' eye-movement patterns or traits. MFD, FTS, and SPG, on the contrary, are valuable indications of the subjects' eye-movement characteristics. Therefore, these measurements will be examined further in Chapter 6.

Chapter 6. RA and STR: A Macro Level Analysis

Eye-tracking data can be analysed at both a macro level and a micro level (Saldanha and O’Brien 2014: 143). Macro-level analysis is understood as one that analyses large units, such as examining the entire text material based on eye-tracking data collected in total. Micro-level analysis, on the other hand, is understood as one that analyses small units, such as one individual fixation on a specific position. This chapter is a macro-level analysis, presenting and discussing data collected on the entire reading material during Reading Aloud (RA) and the entire ST during Sight Translation (STR). Such macro-level analyses have been carried out by many researchers. For example, Pavlović and Jensen (2009) attempted to investigate translation directionality with the help of four eye-tracking data indicators: gaze time, average fixation duration, total task time, and pupil dilation.

All the numerosity measures calculated and analysed in Chapter 6 are dependent variables and precisely quantifiable data. The event detection parameters of the present research are: (a) a minimum peak velocity of 30°/s and (b) a minimum duration of 60 ms for the fixations. Section 6.1 in this chapter carries out both inter- and intra-group comparisons to investigate the investment of cognitive effort in each task. Some well-established eye-tracking indicators are used, including pupil diameter, total task time (TTT), pure gaze activities that is also known as the sum of fixation and saccade duration (FSD), total fixation duration (TFD), and fixation count (FC). Section 6.2 explores the reliability and validity of the mean fixation duration (MFD). Section 6.3 investigates two other potential indications that were used in the eye-tracking data quality assessment, namely Fixation Time on Source text as a percentage of total task time (FTS) and Saccade duration as a Percentage of pure Gaze activity (SPG).

6.1 The Investment of Cognitive Effort

Three groups of subjects took part in the experiments and carried out five different tasks in total. The eye-tracking recordings that passed the data quality assessment are further analysed to examine the Chinese, German, and British subjects' eye-movements in the RA and the STR tasks. The present study employs visual attention as a proxy for cognitive effort, the mental effort involved during RA and STR. Through a preliminary cross-group comparison, three groups' performances in RA-E are explored in Section 6.1.1. Following that, Section 6.1.2 presents the more in-depth intra-group comparisons of the RA and the STR process within Group A and Group C.

The present researcher uses the following five measurements to gain a better understanding of the cognitive process of the subjects in Section 6.1: pupil diameter, TTT, FSD, TFD, and FC. These measurements were calculated and exported by Tobii Studio 3.3.2. The present research aims at testing the following hypotheses in this section:

a. Subjects of different L1s invest different amounts of cognitive effort in processing the same reading material during RA.
b. Subjects invest different amounts of cognitive effort during STR and RA.

6.1.1 A Preliminary Cross-group Comparison of Reading Aloud in English (RA-E)

The present research chose to be cautious when comparing subjects coming from different ethnic backgrounds. Therefore, pupil size, which is essentially a physical feature, is not applied as an indicator when different test groups are compared. The analysis starts with total task time (TTT).

6.1.1.1 Total Task Time (TTT)

Task complexity has been shown to have an impact on the amount of time spent to complete the task (Shreve and Diamond 1997). The use of the Total Task Time (TTT) in some previous studies is based on O'Brien's (2008: 85) argument that "processing speed is a good measurement for cognitive effort on the basis that difficult tasks generally take longer than easier tasks”.

The average TTT results of three test groups during the task Reading Aloud in English (RA-E) are presented in Figure 6.1a.

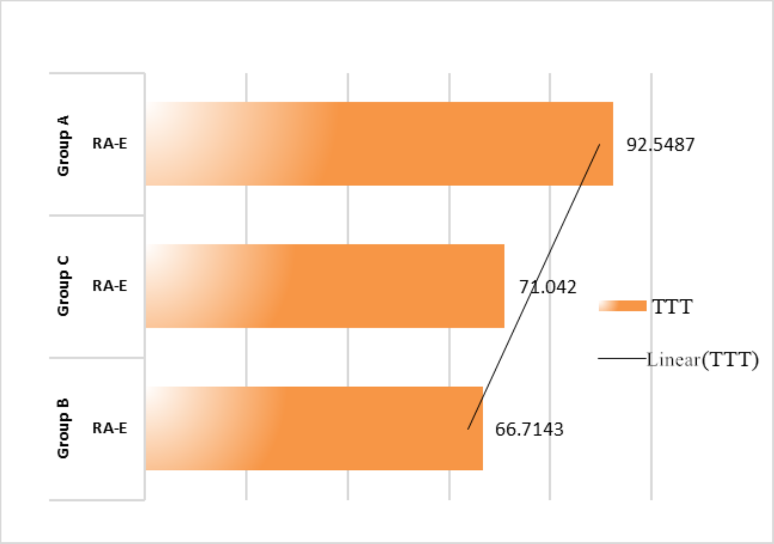


Figure 6.1a: Average TTT (in Seconds) of Group A, B, and C in RA-E

Figure 6.1a displays a linear relation among the three groups’ TTT results. The subjects in Group B have the shortest TTT in the RA-E task. The average TTT of Group A is 92.5487 seconds (n = 30, *SD* = 10.1016), the average FSD of Group B is 66.7143 seconds (n = 35, *SD* = 9.4761), and the average FSD of Group C is 71.042 seconds (n = 20, *SD* = 6.0319). Compared with subjects in Group B, whose L1 is English, subjects in Group C, whose L1 is German, spent on average two seconds more to read the same passage written in English. In contrast, subjects in Group A, whose L1 is Chinese, clearly have a much longer TTT on average.

The first cross-group comparison is made between the TTT of Group A and Group B. Both descriptive statistics and inferential statistics are presented in Table 6.1.1a. Assuming that subjects in Group A have a significantly longer TTT, the *t*-test is one-tailed and unpaired. The Null Hypothesis (H_0) is $\mu 1$ (Group A TTT) $\leqslant$ $\mu 2$ (Group B TTT) (a = 0.05). Detailed results are displayed in Table 6.1.1a* (see Appendix 4). An equal-variance *t*-test is carried out in conjunction with the Cohen’s *d* effect size measurement to test the null hypothesis of zero difference.

Table 6.1.1a: TTT (in Seconds) of Group A and B in RA-E

	TTT	
Group	Group A	Group B
μ	92.5487	66.7143
SD	10.1016	9.4761
p (T$\leqslant$t) one tail	$5.49287E-16 < 0.01$	
Cohen's *d*	$2.64 > 0.8$	
*H_0	$\mu1$ (Group A TTT) $\leqslant$ $\mu2$ (Group B TTT) (a = 0.05)	

Since one variance is not more than 4 times of the other, the *t*-test assuming equal variances is used. Table 6.1.1a* shows that the *t*-statistic (10.6289) is much higher than the one-tail critical *t*-value (1.6694). The corresponding *p*-value is 5.49287E-16 ($p < 0.01$). The significance level here is 0.01. Since *p* is less than 0.01, the hypothesis test is statistically significant. In this case, the corresponding confidence level is 99%. The present research is highly confident that the H_0 can be rejected and that Group B does have a shorter TTT in this task. The Cohen’s *d* measures for this *t*-test, which is the standardised effect size that conveys the size of an effect relative to the variability in the samples, is calculated by the formula:

$$\text{Cohen's } d = (M_2 - M_1) / SD_{\text{pooled}}$$

$$SD_{\text{pooled}} = \surd\,[(SD_1^2 + SD_2^2)/2]$$

A *d* value exceeding 0.2 is considered a *small* effect size; a *d* value exceeding 0.5 is considered a *medium* effect size; a *d* value exceeding 0.8 is considered a *large* effect size. If two datasets’ means do not differ by 0.2 standard deviations or more, the difference is trivial, even though the *t*-test shows that it is statistically significant. For this *t*-test, the *d* value (Cohen's $d = 2.6378$) greatly exceeds the reference point for a large effect (0.8). Hence, based on their TTT, the present study suggests that the British subjects in Group B invested statistically less cognitive effort to finish in the RA-E task.

The second comparison is made between the TTT of Group A and Group C. Table 6.1.1b contains both descriptive statistics and inferential statistics. Details are displayed in Appendices (see Appendix 4, Table 6.1.1b*). The Null Hypothesis of the one-tailed and unpaired *t*-test is $\mu 1$ (Group A TTT) $\leqslant$ $\mu 2$ (Group C TTT) (a = 0.05). To test the null hypothesis of zero difference, an equal-variance t-test is carried out in conjunction with the Cohen's *d* effect size measurement that measures the differences between the datasets.

Table 6.1.1b: TTT (in Seconds) of Group A and C in RA-E

	TTT	
Group	Group A	Group C
μ	92.5487	71.042
SD	10.1016	6.0319
p (T≤t) one tail	1.68551E-11 < 0.01	
Cohen's *d*	2.59 > 0.8	
*H_0	μ1 (Group A TTT)≤μ2 (Group C TTT) (a = 0.05)	

The *t*-statistic (8.5429) shown in Table 6.1.1b* is higher than the one-tail critical *t*-value (1.6772) and the corresponding $p < 0.01$. Hence, there is a very low probability that the sample mean happened by chance. The present research has a high confidence level that the two datasets are significantly different. Therefore, the H_0 is rejected. The present research argues that Group C does have a significantly shorter TTT. The standardised effect size (Cohen's $d = 2.5851$) is considerably more significant than the reference point of a *large* effect (0.8). Hence, based on the TTT results, it is suggested that the subjects in Group C, whose L1 is German, invested a more significant amount of cognitive effort to finish RA-E than the subjects in Group A, whose L1 is Chinese.

As indicated earlier in this section, the difference between Group B's and Group C's TTT is relatively small. In order to see if this difference is statistically significant, a *t*-test assuming equal variances is carried out. The Null Hypothesis is $\mu 1$ (Group C TTT) $\leqslant$ $\mu 2$ (Group B TTT) (a = 0.05). An equal-variance *t*-test is carried out in conjunction with the Cohen's *d* effect size measurements. The descriptive statistics and the results of the *t*-test are presented in Table 6.1.1c. More details are displayed in Appendices (see Appendix 4, Table 6.1.1c*).

Table 6.1.1c: TTT (in Seconds) of Group B and C in RA-E

	TTT	
Group	Group C	Group B
μ	71.042	66.7143
SD	6.0319	9.4761
p (T≤t) one tail	0.0359 < 0.05	
Cohen's *d*	0.5 < 0.54 < 0.8	
*H_0	μ1 (Group C TTT) ≤ μ2 (Group B TTT) (a = 0.05)	

The *t*-statistic (1.8369) is slightly greater than the one-tail critical *t*-value (1.6741). The corresponding *p*-value is 0.0359 ($p < 0.05$). The Cohen's *d*, which is 0.5448 in this case, is considered a medium effect size, as the reference points for the *small*, *medium*, and *large* effects are 0.2, 0.5, and 0.8 respectively. The present study, however, still has an acceptable confidence level to reject the H_0, and to argue that the TTT of Group B is shorter. Therefore, it is assumed that compared to the non-native speakers, the native English speakers devoted less cognitive effort in the RA-E process.

In summary, based on the three groups' sum of Total Task Time (TTT) in the task RA-E, one could hypothesise that the subjects in Group B devoted the least amount of attention and cognitive effort in the RA-E task. The non-native speakers in Group C had to devote more cognitive resources than the native English speakers, and the subjects in Group A devoted the largest amount of cognitive effort in order to complete this task. In the next section, this hierarchy is going to be tested with another indicator named The Sum of Fixation and Saccade Duration (FSD)

6.1.1.2 The Sum of Fixation and Saccade Duration (FSD)

After applying TTT to analyse RA-E, the present research employs the Sum of Fixation and Saccade Duration (FSD) in this section. Essentially, TTT and FSD are on a hierarchy of data purity: TTT is a sum of gaze samples and non-gaze samples; gaze samples contain the pure gaze activities and noises. As mentioned in Chapter 4, pure gaze activity is the sum of the pure fixations and saccades. In the eye-tracking study carried out by Liu, Zheng, and Zhou (2018), this measurement is named FSD (the sum of Fixation and Saccade Duration). To avoid confusion, the present study chooses to use the term FSD in the analysis.

The average FSD of three test groups during the task Reading Aloud in English (RA-E) are presented in Figure 6.1b.

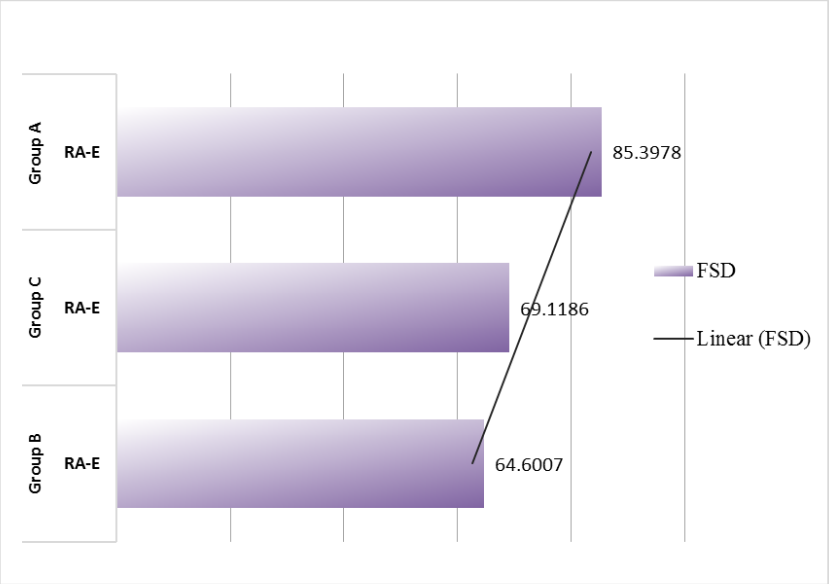


Figure 6.1b: Average FSD (in Seconds) of Group A, B, and C in RA-E

Figure 6.1b displays a linear relation among the three groups' FSD results. The subjects in Group B have the shortest FSD in the RA-E task. The average FSD of Group A is 85.3978 seconds (n = 30, *SD* = 9.6343), the average FSD of Group B is 64.6007 seconds (n = 35, *SD* = 10.2263), and the average FSD of Group C is 69.1186 seconds (n = 20, *SD* = 7.2262). In comparison to Group B and Group C, the Chinese subjects in Group A clearly have a much larger sum of fixation and saccade duration on average. The differences between

Group A and the other two groups are quite noticeable, but the question is whether the difference between Group C and Group B is statistically significant. Therefore, further analyses are carried out.

The first cross-group comparison is made between the FSD of Group A and Group B. Both the descriptive statistics and the inferential statistics are presented in Table 6.1.1d. More details about this *t*-test are displayed in Appendices (see Appendix 4, Table 6.1.1d*). Assuming that the Group A subjects have a significantly longer FSD in the RA-E task, the *t*-test is one-tailed and unpaired. The Null Hypothesis (H_0) is $\mu 1$ (Group A FSD) $\leq \mu 2$ (Group B FSD) (a = 0.05). An equal-variance *t*-test is carried out in conjunction with the Cohen's *d* effect size measurement to test the null hypothesis of zero difference.

Table 6.1.1d: FSD (in Seconds) of Group A and B in RA-E

	FSD	
Group	Group A	Group B
μ	85.3978	64.0067
SD	9.6343	10.2263
p (T≤t) one tail	3.55687E-11 < 0.01	
Cohen's *d*	2.15 > 0.8	
*H_0	μ1 (Group A FSD) ≤ μ2 (Group B FSD) (a = 0.05)	

The *t*-statistic (7.8286) is much higher than the one-tail critical *t*-value (1.6694). The corresponding *p*-value is 3.55687E-11, which is significantly lower than 0.01. Therefore, there is an extremely low probability that the difference between Group A's and Group B's FSD happened randomly. The present research has an exceptionally high confidence level to reject the H_0 and to argue that Group B does have a shorter FSD on the ST during RA-E. For this *t*-test, the *d* value (Cohen's *d* = 2.1531) greatly exceeds the reference point for a large effect (0.8). Hence, the present study suggests that the British subjects in Group B invested statistically less cognitive effort in the RA-E process compared to Chinese subjects in Group A, based on their sum of fixation and saccade duration.

The second comparison is made between the FSD of Group A and Group C. Both the descriptive statistics and the inferential statistics are presented in Table 6.1.1e. More details about this *t*-test are displayed in Appendices (see Appendix 4, Table 6.1.1e*). The Null Hypothesis of the one-tailed and unpaired *t*-test is $\mu 1$ (Group A FSD) ≤ $\mu 2$ (Group C FSD) (a = 0.05). To test the null hypothesis of zero difference, an equal-variance t-test is carried out in conjunction with the Cohen's *d* effect size measurement that measures the differences between the datasets.

Table 6.1.1e: FSD (in Seconds) of Group A and C in RA-E

	FSD	
Group	Group A	Group C
μ	85.3978	69.1186
SD	9.6343	7.2262
p (T≤t) one tail	2.68519E-08 < 0.01	
Cohen's *d*	1.91 > 0.8	
*H_0	μ1(Group A FSD)≤μ2(Group C FSD) (a=0.05)	

A *t*-test assuming equal variances is used based on the variances of the two datasets. The *t*-statistic (6.4371) is higher than the one-tail critical *t*-value (1.6772), and the corresponding $p < 0.01$. Hence, the present research has a high confidence level that the two datasets are significantly different. The H_0 is rejected. The present research argues that Group C does have a significantly shorter FSD during RA-E. The standardised effect size (Cohen's $d = 1.9116$) is considerably more significant than the reference point of a *large* effect (0.8). Hence, based on the FSD results, it is suggested that the native German speakers in Group C invested a more significant amount of cognitive effort during RA-E than the native Chinese speakers in Group A.

As indicated earlier in this section, compared with the native English speakers in Group B, the German subjects in Group C had a longer FSD on average in the RA-E task. The difference is relatively small. In order to see if this difference is statistically significant, a *t*-test assuming equal variances is carried out. The Null Hypothesis is $\mu 1$ (Group C FSD) ≤ $\mu 2$ (Group B FSD) (a = 0.05). An equal-variance *t*-test is carried out in conjunction with the Cohen's *d* effect size measurements that measure the differences between the datasets to test the null hypothesis of zero difference. The descriptive statistics and the results of the *t*-test are presented in Table 6.1.1f. Table 6.1.1f* in Appendices (see Appendix 4) contains more details about this *t*-test.

Table 6.1.1f: FSD (in Seconds) of Group B and C in RA-E

	FSD	
Group	Group C	Group B
μ	69.1186	64.0067
SD	7.2262	10.2263
p (T≤t) one tail	0.043836 < 0.05	
Cohen's *d*	0.2 < 0.45 < 0.5	
*H_0	μ1(Group C FSD)≤μ2(Group B FSD) (a=0.05)	

The *t*-statistic (1.7399) in this *t*-test is slightly greater than the one-tail critical *t*-value (1.6741). The corresponding *p*-value is 0.0438 ($p < 0.05$). The Cohen's *d*, which is 0.4479 in this case, is considered a close-to-medium effect size. The present study, however, still has an acceptable confidence level to reject the H_0, and to argue that the British subjects in Group B had a shorter FSD on the ST during RA-E. Therefore, it is assumed that compared to the non-native speakers, the native English speakers devoted less cognitive effort in the RA-E process.

In summary, based on the three groups' sum of Fixation and Saccade Durations (FSD) in the task RA-E, one could further hypothesise that the subjects in Group B devoted the least amount of attention and cognitive effort in the RA-E task. The non-native speakers in Group C had to devote more cognitive resources than the native English speakers, and the subjects in Group A devoted the largest amount of cognitive effort in order to complete this task. In the next section, this hierarchy is going to be tested with another indicator: total fixation duration (TFD).

6.1.1.3 Total Fixation Duration (TFD)

Total fixation duration is commonly used to index cognitive effort in translation studies and STR studies (e.g., Pavlović and Jensen 2009; Korpal 2012; Hvelplund 2014). Total fixation duration is a sum of all fixations, calculated by "combining fixation duration and number of fixations" (Hogaboam 1983: 310). TFD is said to be "the most used measure in eye-tracking research" (Holmqvist *et al*. 2011: 377). It is acknowledged by many researchers that a longer total fixation duration and more frequent visits to the AOI indicate a greater cognitive effort devoted to the task (Saldanha and O'Brien 2014: 144).

Each test group's average total fixation durations (TFD) in the task RA-E are calculated by Tobii Studio 3.3.2 and presented in Figure 6.1c.

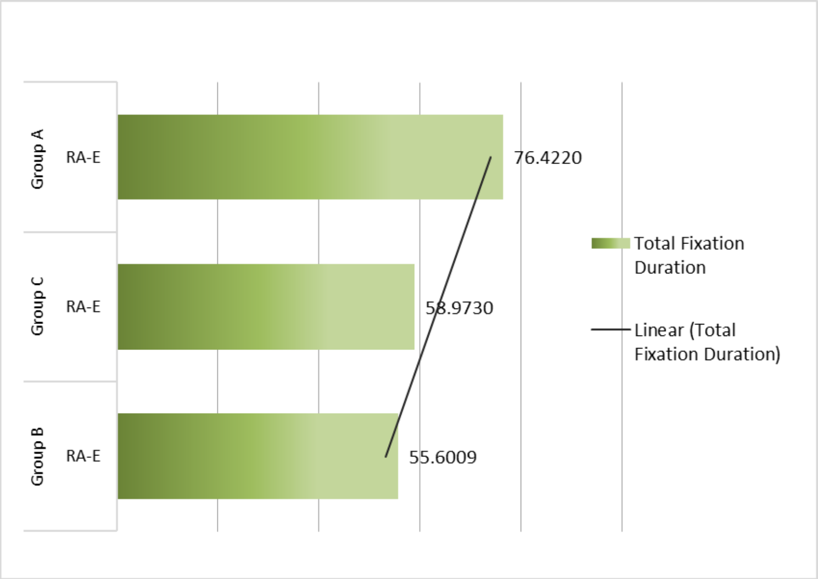


Figure 6.1c: Average TFD (in Seconds) of Group A, B, and C in RA-E

It is shown in Figure 6.1c that, on average, the British subjects have the shortest total fixation duration in the RA-E task. The average TFD of Group A is 76.4220 seconds (n = 30, *SD* = 9.8171); the average TFD of Group B is 55.6009 seconds (n = 35, *SD* = 7.8701); and the average TFD of Group C is 58.9730 seconds (n = 20, *SD* = 5.9451) in the RA-E task. Among the three test groups, Group A clearly has the longest total fixation duration on average.

The differences between Group A and the other two groups are large, whereas the difference between Group C and Group B is considerably smaller. Therefore, further analyses are carried out to see whether these differences are statistically significant, or whether they could have reasonably occurred randomly.

First of all, the total fixation duration of Group A and Group B are contrasted. Assuming that the subjects in Group A have a significantly longer total fixation duration on the reading material in the RA-E task, the *t*-test ought to be one-tailed and unpaired. The Null Hypothesis is $\mu 1$ (Group A TFD) $\leq \mu 2$ (Group B TFD) (a = 0.05). To test the null hypothesis of zero difference between Group A's and Group B's TFD, an equal-variance *t*-test is carried out in conjunction with the Cohen's *d* effect size measurement. Both the descriptive and the inferential statistics are presented in Table 6.1.1g. Table 6.1.1g* in Appendices (see Appendix 4) contains more details about this *t*-test.

Table 6.1.1g: TFD (in Seconds) of Group A and B in RA-E

	TFD	
Group	Group A	Group B
μ	76.4220	55.6009
SD	9.8170	7.8701
p (T≤t) one tail	1.04875E-12 < 0.01	
Cohen's d	2.16 > 0.8	
*H_0	μ1(Group A TFD)≤μ2(Group B TFD) (a=0.05)	

Group A's dataset is significantly different from that of Group B's. The *t*-statistic (8.7056) of this *t*-test is higher than the one-tail critical *t*-value (1.6694). The corresponding *p*-value is 1.04875E-12, which is significantly lower than 0.01. Therefore, the H_0 is rejected with a high confidence level. Compared to Group A, Group B does have a significantly shorter total fixation duration on the ST during RA-E. Furthermore, the *d* value (Cohen's $d = 2.1598$) that conveys the size of an effect relative to the variability in the samples greatly exceeds the reference point of a *large* effect size (Cohen's *d*

= 0.8). As a result, it is found that these native English speakers used less cognitive effort in the RA-E task.

The second comparison is made between the TFD of Group A and Group C. Assuming that the subjects in Group A have a significantly longer total fixation duration than the subjects in Group C in the RA-E task, the *t*-test should be one-tailed and unpaired. The Null Hypothesis is $\mu 1$ (Group A TFD) $\leq \mu 2$ (Group C TFD) ($a = 0.05$). To test the null hypothesis of zero difference, an equal-variance t-test is carried out in conjunction with the Cohen's *d* effect size measurement. The results are presented in Table 6.1.1h. Table 6.1.1h* in Appendices (see Appendix 4) contains more details about this *t*-test.

Table 6.1.1h: TFD (in Seconds) of Group A and C in RA-E

	TFD	
Group	Group A	Group C
μ	76.4220	58.9730
SD	9.8170	5.9451
p (T≤t) one tail	2.42926E-09 < 0.01	
Cohen's *d*	2.15 > 0.8	
*H0	μ1(Group A TFD)≤μ2(Group C TFD) (a=0.05)	

Table 6.1.1h* in Appendices (see Appendix 4) shows that the *t*-statistic (7.1128) is greater than the one-tail critical *t*-value (1.6772). The significant difference is found to be at the $p < 0.01$ level. As a result, there is an extremely low probability that the data occurred randomly. Hence, the H_0 is rejected with a high confidence level. In summary, Group C's total fixation duration on the ST during RA-E is shorter than that of Group A's. The effect size measure of this *t*-test (Cohen's *d*) is 2.1501, which is considerably larger than a *large* effect size (a Cohen's *d* of 0.8). Therefore, it is found that compared to the native Chinese speakers, the native German speakers devoted less cognitive effort during the RA-E process.

The third comparison is made between the TFD of Group B and Group C. Compared with the native English speakers in Group B, Group C subjects' total fixation duration is 3.37 seconds longer on average. Indeed, the difference is relatively small. A *t*-test assuming equal variances is carried out. The Null Hypothesis is $\mu 1$ (Group C TFD) ≤ $\mu 2$ (Group B TFD) (a = 0.05) to check whether this difference is statistically significant. An equal-variance t-test is carried out in conjunction with the Cohen's *d* effect size measurement that measures the differences between the datasets. Both the descriptive and the inferential statistics are presented in Table 6.1.1i (see Appendix 4, Table 6.1.1i* for more details about this *t*-test).

Table 6.1.1i: TFD (in Seconds) of Group B and C in RA-E

	TFD	
Group	Group C	Group B
μ	58.9730	55.6009
SD	5.9451	7.8701
p (T≤t) one tail	0.051224 > 0.05	
Cohen's *d*	0.2 < 0.39 < 0.5	
*H_0	μ1(Group C TFD)≤μ2(Group B TFD) (a=0.05)	

The one-tail critical *t*-value (1.6741) is slightly higher than the *t*-statistic (1.6618). The corresponding *p*-value is 0.05122, which is just a bit higher than 0.05. As a result, the present research could not reject the H_0 with a very high confidence level in this case. Furthermore, the Cohen's *d* of 0.3931 is considered a *small* effect size, meaning that the difference between the two datasets is relatively trivial. The inferential statistics cannot corroborate the descriptive statistics. Instead, it is thus hypothesised that there is no significant difference between Group B's and Group C's total fixation duration on the ST in the RA-E task. Based on this analysis, it appears that the subjects in Group C and Group B might have invested approximately the same amount of cognitive effort during RA-E.

Based on the results of the three groups' TFD in the task RA-E, one could further hypothesise that the Chinese subjects in Group A devoted the largest amount cognitive effort, while the German subjects in Group C and the British subjects in Group B might have invested approximately the same amount of cognitive effort in RA-E. Nevertheless, this hierarchy does not contradict the one formed by FSD. It is thus hoped that another indicator would be able to present the difference between Group B and C in a clearer way.

6.1.1.4 Fixation Count (FC)

Two eye-movement indicators, namely fixation count and mean fixation duration, seem to have a rather complicated relationship with each other. The researchers using eye-tracking data in reading process research often stated that the *average fixation duration* and the *fixation count* are both indexes of cognitive effort. Hvelplund (2014: 212) summarised it as follows: "Longer fixations and more fixations indicate more effortful processing and shorter fixations and fewer fixations indicate less effortful processing, and this more effortful processing is often linked to an increase in difficulty". However, it should be emphasised that the *fixation duration* here does not equal the *mean fixation duration*. The former is a sum of all the fixations whereas the latter is calculated by dividing the total fixation duration by the number of fixations.

Having a larger number of fixations and having a longer average fixation duration are somehow contradictory with each other. When the trial duration is fixed, there should be a clear negative correlation between the number of fixations and the mean duration of fixations: the more fixations, the shorter the MFD, and vice versa (Holmqvist *et al*. 2011: 412). It, therefore, could be possible that readers would trade off the number of fixations for the mean fixation duration, or vice versa (Hogaboam 1983: 310).

Some researchers have voiced their concern regarding the relationship between the task complexity and these eye movement indicators. For instance, in an eye-tracking based study conducted by Sharmin *et al*. (2008), the student subjects cast a larger number of fixations on the more complicated ST than the relatively easier ST, but the total fixation durations were almost the same. Moreover, in the study carried out by Jakobsen and Jensen (2008), the average MFD did not show a significant difference between two different reading tasks and a STR task. Especially, reading for general comprehension and reading for translation yielded precisely the same MFD result (205 ms) in their experiment. In these cases, the subjects taking part in the experiments might have traded off the length of the mean fixation duration for the number of fixations.

During RA, it is possible that the information "falling in the center of vision on fixation *n* received some preliminary processing on fixation *n*-1 when they were in parafoveal vision" (Rayner 1983: 106). Since one's reading competence influences the size of the perceptual span (Rayner 1995: 7), it has an impact on fixation count as well. This is because when one's perceptual span is relatively long, he/she would need a smaller number of fixations to cover the entire text. However, when the cognitive demand increases, it reduces the ease of extending one's perceptual span, resulting in an increase in the number of fixations.

To summarise, a more substantial number of fixations is associated with a larger investment of cognitive effort. The number of fixations has been used as an operational definition of many things, including the search efficiency and difficulty in interpreting and the reading depth (Holmqvist *et al*. 2011: 412-413). Fixation count has also been applied by Korpal (2012) in an eye-tracking based STR research to indicate cognitive effort. Based on the aforementioned literature, the present research chooses to employ fixation count as one of the indicators to investigate the investment of cognitive effort.

Three group's fixation counts (FC) during reading aloud in English (RA-E) are presented in Figure 6.1d.

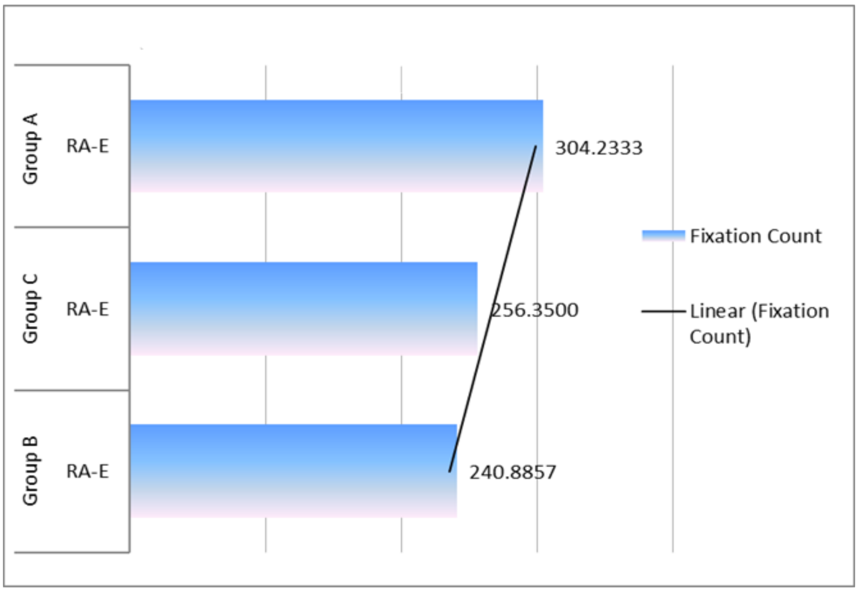


Figure 6.1d: Average FC (in Numbers) of Group A, B, and C in RA-E

Figure 6.1d presents a clear regularity. The British subjects devoted the smallest number of fixations to complete the RA-E task. Compared to the native English speakers, the German subjects spent about 15.46 more fixations on average to read the same English passage. In contrast, among the three test groups, the Chinese subjects in Group A clearly had many more fixations on average than the speakers of the other two European languages. The average fixation count of Group A is 304.2333 (n = 30, *SD* = 32.3085); the average fixation count of Group B is 240.8857 (n = 35, *SD* = 36.8612); and the average fixation count of Group C is 256.3500 (n = 20, *SD* = 22.6606) in the task RA-E. The differences between Group A and the other two groups are large, whereas the difference between Group C and Group B is considerably smaller. Therefore, further analyses are carried out to help determine whether these differences are statistically significant.

First of all, Group A's and Group B's fixation count (FC) in RA-E are compared. Assuming that the subjects in Group A have a significantly larger number of fixations on the reading material in the RA-E task, the *t*-test should be one-tailed and unpaired. The Null Hypothesis is $\mu 1$ (Group A FC) $\leq \mu 2$ (Group B FC) (a = 0.05). To test the null hypothesis of zero difference, an equal-variance *t*-test is carried out in conjunction with the Cohen's *d* effect size measurement, which measures the differences between the datasets. Both the descriptive and the inferential statistics are presented in Table 6.1.1j. More details about this *t*-test are displayed in Appendices (see Appendix 4, Table 6.1.1j*).

Table 6.1.1j: FC (in Numbers) of Group A and B in RA-E

	FC	
Group	Group A	Group B
μ	304.2333	240.8857
SD	32.3085	36.8612
p (T≤t) one tail	2.89813E-10 < 0.01	
Cohen's *d*	1.83 > 0.8	
*H_0	μ1(Group A FC)≤μ2(Group B FC) (a=0.05)	

The *t*-statistic (7.3079) turns out to be greater than the one-tail critical *t*-value (1.6694). The corresponding *p*-value is 2.89813E-10, which is significantly lower than 0.01. Hence, there is an extremely low probability that the data are just random. Group B does have a noticeably smaller number of fixations on the ST during RA-E than Group A. The present research, therefore, can confidently reject the H_0. The *d* value (Cohen's $d = 1.8277$), which conveys the size of an effect relative to the variability, in the samples reflects a *large* effect size. Based on the findings here, one could suggest that the native English speakers devoted less cognitive effort in the RA-E process.

Then, the second comparison between Group A and Group C is carried out. Both the descriptive statistics and the inferential statistics are shown in Table 6.1.1k. A *t*-test assuming unequal variances is carried out due to the unequal variances, and its results are presented in Table 6.1.1k* (See Appendix 4). The Null Hypothesis is $\mu 1$ (Group A FC) ≤ $\mu 2$ (Group C FC) (a = 0.05). The *t*-test is carried out in conjunction with the Cohen's *d* effect size measurement that measures the differences between the datasets.

Table 6.1.1k: FC (in Numbers) of Group A and C in RA-E

	FC	
Group	Group A	Group C
μ	304.2333	256.35
SD	32.3085	22.6606
p (T≤t) one tail	3.07729E-07 < 0.01	
Cohen's *d*	1.72 > 0.8	
*H0	µ1(Group A FC)≤µ2(Group C FC) (a=0.05)	

The *t*-statistic (5.7439) is higher than the one-tail critical *t*-value (1.6772). The corresponding *p*-value is 3.07729E-07, which is significantly low. Hence, there is a high level of confidence to reject the H_0 and to argue that the German subjects in Group C cast a noticeably smaller number of fixations on the ST during RA-E. The significant difference is found to be at the $p < 0.01$ level. Furthermore, the effect size measure that conveys the size of an effect relative to the variability in the samples is calculated. In this case, the Cohen's *d* of 1.7159 suggests that it is a very *large* effect size. Therefore, compared to the native Chinese speakers, the native German speakers devoted less cognitive effort in the RA-E process.

Finally, the last comparison is made between Group B and Group C. Based on the descriptive statistics, it is assumed that the subjects in Group C cast a significantly larger number of fixations than the subjects in Group B in the RA-E task. The one-tailed and unpaired *t*-test is carried out in conjunction with Cohen’s *d* effect size measurement. The Null Hypothesis is $\mu 1$ (Group C FC) ≤ $\mu 2$ (Group B FC) (a = 0.05). The descriptive and inferential results are presented in Table 6.1.1l. More details are presented in Table 6.1.1l* (See Appendix 4).

Table 6.1.1l: FC (in Numbers) of Group B and C in RA-E

	FC	
Group	Group B	Group C
μ	240.8857	256.35
SD	36.8612	22.6606
p (T≤t) one tail	0.047693< 0.01	
Cohen's *d*	0.5 < 0.51 < 0.8	
*H0	μ1(Group C FC)≤μ2(Group B FC) (a=0.05)	

Table 6.1.1l* contians the *t*-statistic (1.6979), which is greater than the one-tail critical *t*-value (1.6741). The *p*-value is 0.047692974. The present research has an acceptable confidence level to reject the H_0 and to argue that Group C does have a noticeably larger number of fixations on the ST during RA-E. In this case, the *d* value (Cohen's $d = 0.5054$) is close to a *medium* effect size. Since fixation count has a negative correlation with one's perceptual vision span, and since "part of the skill of reading is using the parafoveal information to encode words rapidly enough" (Rayner and Pollatsek 1989: 164), fixation count shows how efficient the encoding of the words is. As the subjects in Group B are more efficient in encoding the parafoveal information, they are able to finish the reading task with a smaller number of fixations. Based on the comparison of these two groups' fixation counts, the present study suggests that the native English speakers devoted less cognitive effort in the RA-E process than the German subjects.

In summary, the non-native speakers have more fixations on the reading material during RA because they are less efficient in using the parafoveal and the peripheral information (Rayner and Pollatsek 1989: 388). Based on the results of the fixation count, one could further hypothesise that the subjects in Group B devoted the least amount of attention and cognitive effort during RA-E, and the Chinese subjects in Group A invested the most cognitive effort in the task. Unlike total fixation duration, this indicator succeeded in presenting the difference between Group C and Group B. In short, non-native speakers in Group A and Group C devoted more cognitive resources to complete this RA-E task.

In conclusion, TTT, FSD, and FC yield the same result in terms of which group invested the least or the largest amount of cognitive effort in the RA-E task. TFD, on the other hand, does not allow the present study to state that the two datasets of Group C and Group B are significantly different. Nevertheless, the result does not contradict the hierarchy formed with the result of the other two eye-tracking measurements. Instead, it is hypothesised that Group C's cognition investment in the RA-E task is slightly higher than that of Group B's, although the difference between them is not so significant. A hierarchy of the three groups' investment of cognitive effort during RA-E is formed: Group A > Group C ≥ Group B. Therefore, the first hypothesis is confirmed: subjects with different L1s invest different amounts of cognitive effort in processing the same reading material during RA.

6.1.2 An In-depth Intra-group Comparison of RA and STR

In this section, the present study carries out more in-depth intra-group comparisons within Group A and C, aiming to examine the second hypothesis: both Chinese subjects and German subjects invest more cognitive effort in STR tasks than in RA tasks. The following eye-tracking measurements are used as indicators of the amount of attention and cognitive effort invested to complete the tasks: pupil diameter, TTT, FSD, TFD, and FC.

Specifically, Group A's RA-E is contrasted with Group A's STR E-C, Group A's RA-C is contrasted with Group A's STR C-E, and Group C's RA-E is contrasted with Group C's STR E-G. Thus, the comparison is made within individual test groups with the same reading material in both the RA and the STR tasks.

6.1.2.1 Pupil Diameter

Pupil diameter has been used as an important indicator by many researchers who conducted process-oriented studies using an eye-tracking methodology. Nevertheless, researchers have been reminded that "caution should be exercised when collecting and analysing pupil size data, since this type of eye movement is sensitive to not only changes in cognitive load but to many other factors" (Hvelplund 2014: 214). In order to minimise the impact of the changes in light intensity, the experiments took place in the eye-tracking laboratory which has only one stable light source. Also, subjects were asked to confirm that they had not ingested any medicine, caffeine, or alcohol on the day.

In addition, this research is cautious with comparing the pupil size or pupil dilation across different groups because the participants were from different ethnic origins. Instead, the present research looks at individuals and compares each subject's pupil diameter in one task with his/her own pupil diameter in another task. It is believed that the change of one's average pupil size in one task and another gives us information about the amount of his/her cognition allocation.

The Tobii TX300 eye-tracker can record an individual's left and right pupil diameter for each gaze event. The researcher calculated the average left-right pupil diameter and presented the results in Figure 6.1e and Figure 6.1f. Group C's data are analysed first because Group C has completed fewer tasks, and this group consists of fewer subjects. The present study hopes to form an initial hypothesis with Group C's data, and then test it with Group A's dataset, which is considerably larger.

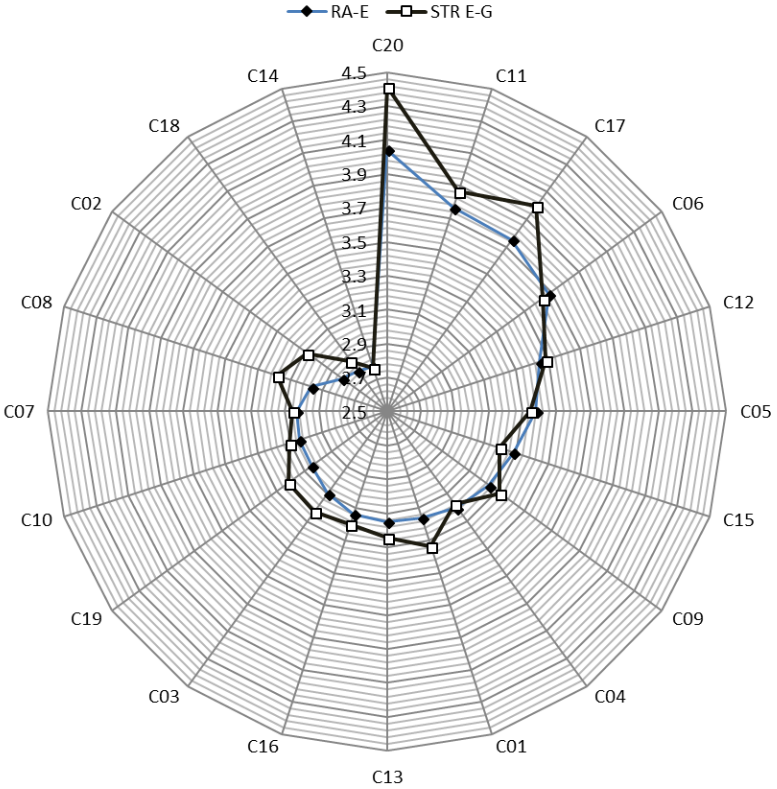


Figure 6.1e: Pupil Size (in Millimetres) of Group C in RA-E and STR E-G

On average, Group C's left-right average pupil diameter is 3.2414 millimetres (mm) (n = 20, *SD* = 0.3446) during RA-E and 3.3361 mm (n = 20, *SD* = 0.3871) during STR E-G.

The majority (75%) of them have larger pupils during STR E-G, with five exceptions. Then, a paired two-sample *t*-test is carried out. The Null Hypothesis is $\mu 1$ (Pupil Size in STR) ≤ $\mu 2$ (Pupil Size in RA) (a = 0.05). Both the descriptive and the inferential are shown in Table 6.1.2a. Table 6.1.2a* in Appendices (see Appendix 4) contains more details about this *t*-test.

Table 6.1.2a: Pupil Size (in Millimetres) of Group C in RA-E and STR E-G

Group C	Pupil Diameter	
Task	RA-E	STR E-G
μ	3.2414	3.3360
SD	0.3446	0.3871
p (T≤t) one tail	0.000846 < 0.01	
Cohen's *d*	0.2 < 0.26 < 0.5	
*H0	μ1(Pupil Size in STR) ≤ μ2(Pupil Size in RA)(a=0.05)	

The t-statistic (3.6526) is much higher than the one-tail critical t-value (1.7291). The significant difference is found to be at the $p < 0.01$ level. Therefore, the present research has a high confidence level to reject the H_0 and to conclude that Group C's pupil size in the STR task is significantly and statistically larger than that in the RA task. The d value (Cohen's $d = 0.2583$) that conveys the size of an effect relative to the variability in the samples is a *small* effect size. The findings indicate that the subjects in Group C invested more cognition resources in the STR task.

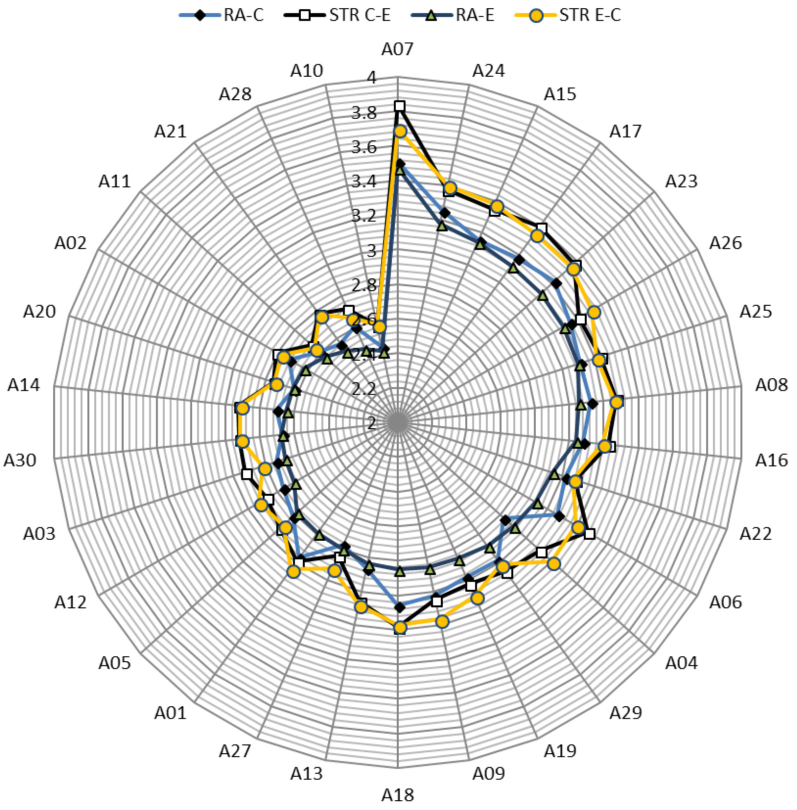


Figure 6.1f: Pupil Size (in Millimetres) of Group A in RA and STR

On average, Group A's left-right average pupil diameter is 2.9204 mm (n = 30, *SD* = 0.2471) during RA-C, 2.8552 mm (n = 30, *SD* = 0.2424) during RA-E, 3.0606 mm (n = 30, *SD* = 0.2685) during STR C-E, and 3.0589 mm (n = 30, *SD* = 0.2621) during STR E-C.

In Group A, every individual's average pupil diameter during STR E-C exceeds their own pupil diameter during RA-E; every individual's average pupil diameter during STR C-E exceeds their own pupil diameter during RA-C. Overall, it indicates that for the Chinese subjects, the two STR tasks are more effort-consuming than the two corresponding RA tasks. Then, two paired two-sample *t*-tests are carried out. The Null Hypothesis is $\mu 1$ (Pupil Size in STR) $\leq \mu 2$ (Pupil Size in RA) (a = 0.05). Both the descriptive and the inferential are shown in Table 6.1.2b. Table 6.1.2b* in Appendices (see Appendix 4) contains more details about this *t*-test.

Table 6.1.2b: Pupil Size (in Millimetres) of Group A in RA and STR

Group A	Pupil Diameter			
Task	STR E-C	RA-E	STR C-E	RA-C
μ	3.0589	2.8552	3.0606	2.9204
SD	0.2621	0.2424	0.2685	0.2471
p (T≤t) one tail	1.05403E-16 < 0.01		3.39634E-11 < 0.01	
Cohen's *d*	0.81 > 0.8		0.5 < 0.54 < 0.8	
*H_0	$\mu 1$(Pupil Size in STR) $\leq \mu 2$(Pupil Size in RA)(a=0.05)			

In the one-tail paired two-sample *t*-test, the *p*-values far lower than 0.01 show that the average results are representative. The present research has an exceptionally high confidence level to reject the H_0 and to argue that Group A's average pupil sizes in the two STR tasks are significantly and statistically larger than the results in the two RA tasks. Furthermore, the effect size measures of the two *t*-tests are calculated. The standardised effect size of the first *t*-test (Cohen's *d* = 0.8069) exceeds the reference point of a *large* effect (0.8) and the *d* value (Cohen's *d* = 0.5434) of the second *t*-test is considered a *medium* effect size.

To sum up, in this intra-group comparison of the RA and the STR process, the amount of Group A's and Group C's cognitive effort is displayed and contrasted within the test groups. As the one indicator that has a less direct correlation with the other eye-movement measurements, the pupil diameter confirms that the subjects in Group A invested a more substantial amount of cognition resources in the two STR tasks.

6.1.2.2 Total Task Time (TTT)

As mentioned in Section 6.1.1, the higher the TTT result, the more cognition has been invested to complete the task.

Group C completed RA-E and STR E-G. Group C's average TTT is 71.0420 seconds (n = 20, *SD* = 6.0319) in the RA task and 165.8745 seconds (n = 20, *SD* = 17.3430) in the STR task. The individuals' TTT results from the two tasks are presented in Figure 6.1g. The triangles represent individuals' TTT in RA-E and the circles represent the individuals' TTT in STR E-G.

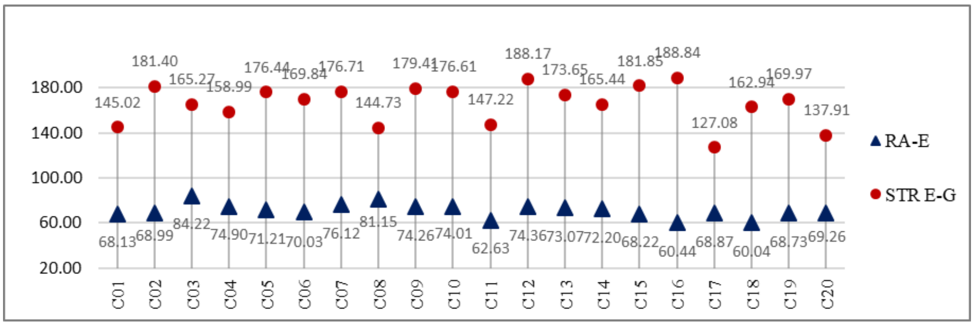


Figure 6.1g: TTT (in Seconds) of Group C in RA-E and STR E-G

Some quantitative differences are shown by the sample. On average, the difference between one German subject's TTT in the RA-E task and the STR E-G task is 94.8325 seconds. Since every individual in Group C has a longer TTT in the STR task, it is evident that the amount of cognitive effort invested in sight translating is more than that invested in reading aloud.

Thereafter, Group C's TTT in the RA-E task and the STR E-C task are investigated. Group A's average TTT is 92.5487 seconds (n = 30, *SD* = 10.1016) in the RA-E task and 156.258 seconds (n = 30, *SD* = 21.5574) in the STR E-C task. The individuals' TTT in the two tasks are presented in Figure 6.1h. The triangles represent the individuals' TTT in RA-E and the circles represent the individuals' TTT in STR E-C.

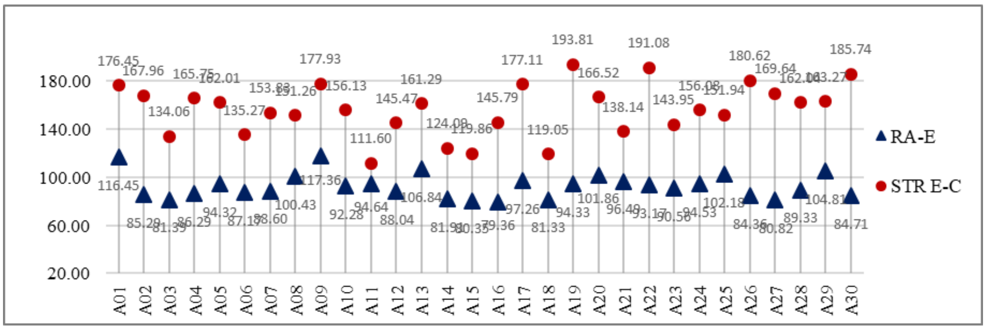


Figure 6.1h: TTT (in Seconds) of Group A in RA-E and STR E-C

The large difference between the two tasks is shown on the chart. On average, the difference between one Chinese subject's FSD in the RA-E task and the STR E-C task is 63.7093 seconds. The chart shows that every Chinese subject has a higher TTT result in the STR E-C task. Therefore, the amount of cognitive effort invested to sight translate the English passage into Chinese is more than that invested to read the same piece of English passage aloud.

Subsequently, Group A's TTT in the task RA-C and the task STR C-E are investigated. Group A's average TTT is 90.4640 seconds (n = 30, *SD* = 8.6041) in the task RA-C and 162.8560 seconds (n = 30, *SD* = 15.8338) in the task STR C-E. The TTT results are presented in Figure 6.1i. The triangles represent individuals' TTT in RA-C and the circles represent individuals' TTT in STR C-E.

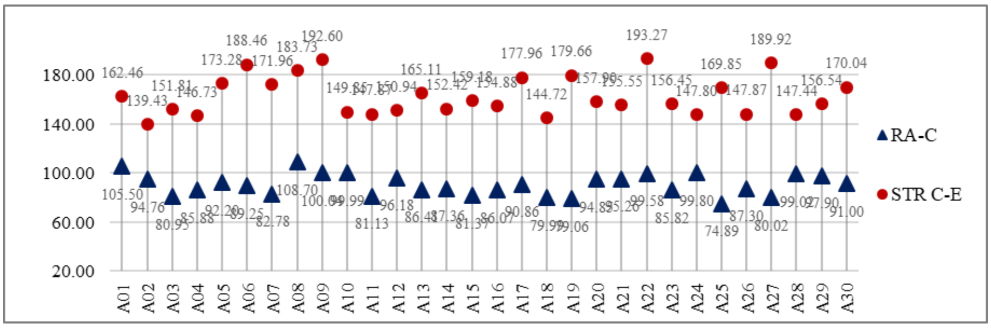


Figure 6.1i: TTT (in Seconds) of Group A in RA-C and STR C-E

On average, the difference between one Chinese subject's TTT in the RA-C task and the STR C-E task is 72.3920 seconds. The individuals' TTT result in the task STR C-E is always higher than that in the task RA-C, which means the amount of the cognitive effort invested in sight translating the Chinese to English is more than that invested in reading aloud the same Chinese text.

Based on the descriptive statistics presented above, it is hypothesised that both STR E-C and STR C-E are much more effort-consuming than the corresponding RA tasks. In addition, three paired-sample *t*-tests are conducted to corroborate this hypothesis. The Null Hypothesis is $\mu 1$ (TTT in STR) $\leqslant$ $\mu 2$ (TTT in RA) (a = 0.05). The results are shown in Table 6.1.2c. More details about this *t*-test are displayed in Appendices (see Appendix 4, Table 6.1.2c*).

Table 6.1.2c: TTT (in Seconds) of Group A and C in RA and STR

	Group A TTT				Group C TTT	
Task	RA-C	STR C-E	RA-E	STE E-C	RA-E	STR E-G
μ	90.4640	162.8560	92.5487	156.2580	71.0420	165.8745
SD	8.6041	15.8338	10.1016	21.5574	6.0319	17.3430
p (T⩽t) one tail	2.54682E-20 < 0.01		4.58792E-17 < 0.01		1.070962E-15 < 0.01	
Cohen's *d*	5.68 > 0.8		3.78 > 0.8		7.30 > 0.8	
*H_0	μ1 (Group A TTT)⩽μ2 (Group B TTT) (a = 0.05)					

The *t*-statistics are all higher than the one-tail critical *t*-values (22.7120 > 1.6991; 17.2132 > 1.6991; 23.1814 > 17.291) (See Appendix 4, Table 6.1.2c.*). The corresponding *p*-values are all well below 0.01. Therefore, this study rejects the H_0 with a high confidence level and argues that both groups' TTT results are significantly different in the STR and the RA tasks. Furthermore, the three *d* values (Cohen's *d*) that convey the size of an effect relative to the variability in the samples are calculated respectively. They (Cohen's *d* = 5.6812, 3.7846, 7.3039) all greatly exceed the reference point of a *large* effect size (Cohen's *d* = 0.8). Therefore, it is found that the subjects in the two experimenal groups devoted less cognitive effort in the RA tasks than in the STR tasks.

In summary, in this intra-group comparison of the RA and the STR process, the amount of Group A's and Group C's cognitive effort was displayed and contrasted within the test groups with the help of the indicator TTT. It is observed that both the Chinese and the German subjects in this experiment have devoted more cognitive effort to the STR tasks than to the RA tasks. This finding is in line with the conclusion obtained from the pupil diameter and it is checked by the next indicator: the sum of fixation and saccade duration (FSD).

6.1.2.3 The Sum of Fixation and Saccade Duration (FSD)

As mentioned in Section 6.1.1, the total duration of the pure gaze activities is the sum of fixation and saccade duration. It has been proposed that the higher the FSD result, the more cognition has been invested to complete the task (Liu, Zheng, and Zhou 2018).

Group C completed two tasks. Group C's average FSD is 69.1186 seconds (n = 20, *SD* = 7.2262) in the RA-E task and 144.5651 seconds (n = 20, *SD* = 18.3574) in the STR E-G task. The individuals' FSD results in the two tasks are presented in Figure 6.1j. The triangles represent individuals' FSD in RA-E, and the rhombuses represent the individuals' FSD in STR E-G.

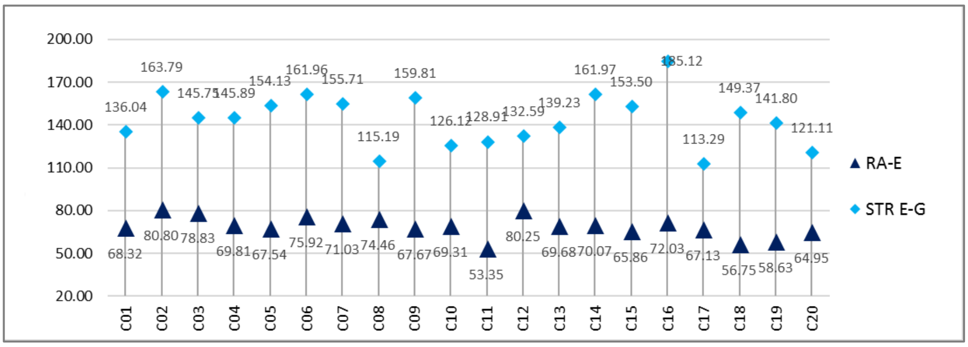


Figure 6.1j: FSD (in Seconds) of Group C in RA-E and STR C-G

The remarkable difference between the two tasks is shown on the chart. On average, the difference between one German subject's FSD in the RA-E task and the STR E-G task is 75.4465 seconds. Since every individual in Group C has a longer FSD in the STR task, it is evident that the amount of attention and cognitive effort invested in sight translating the English passage into German is more than that invested in reading the same English passage aloud.

Thereafter, Group A's FSD in the RA-E task and the STR L2-L1 task are investigated. Group A's average FSD is 85.3978 seconds (n = 30, *SD* = 9.6343) in the RA-E task and 135.6211 seconds (n = 30, *SD* = 18.9142) in the STR E-C task. Individuals' FSD in the two tasks are presented in Figure 6.1k. The triangles represent individuals' FSD in RA-E, and the rhombuses represent the individuals' FSD in STR E-C.

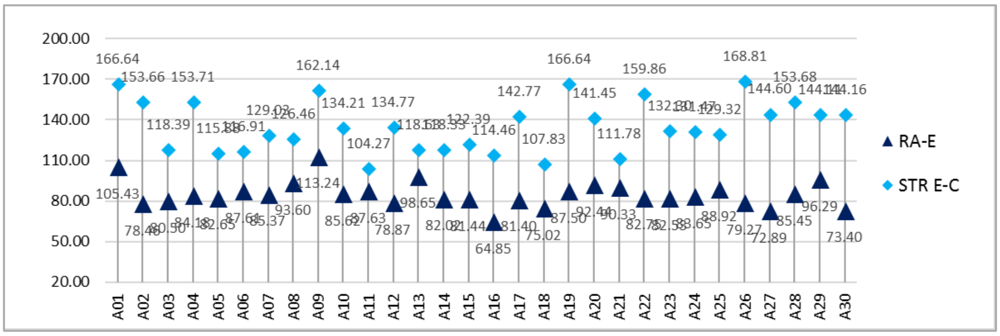


Figure 6.1k: FSD (in Seconds) of Group A in RA-E and STR E-C

Some quantitative differences are shown in the sample: on average, the difference between one Chinese subject's FSD in the RA-E task and the STR E-C task is 50.2334 seconds. The chart shows that every Chinese subject has a higher FSD result in the STR E-C task. Hence, it is suggested that the amount of attention and cognitive effort invested to sight translate the English passage into Chinese is more than that invested to read the same piece of English passage aloud.

Finally, Group A's FSD in the task RA-C and the task STR L1-L2 are investigated. Group A's average FSD is 83.1721 seconds (n = 30, *SD* = 10.8251) in the task RA-C and 138.4341 seconds (n = 30, *SD* = 14.6508) in the task STR C-E. The FSD results are presented in Figure 6.1l. The triangles represent individuals' FSD in RA-C and the rhombuses represent individuals' FSD in STR C-E.

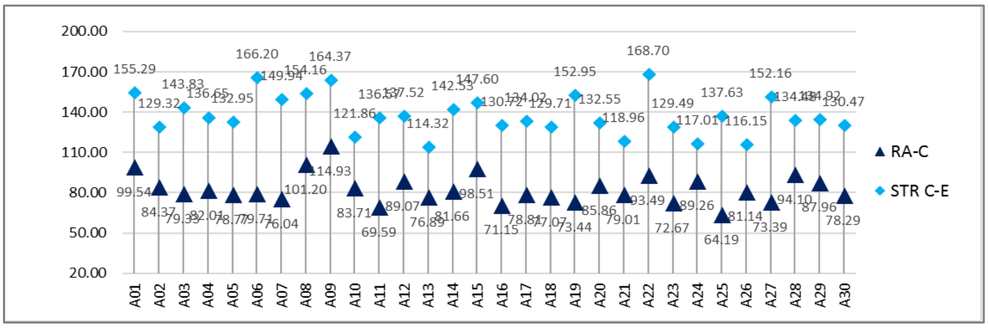


Figure 6.1l: FSD (in Seconds) of Group A in RA-C and STR C-E

On average, the difference between one Chinese subject's FSD in the RA-C task and the STR C-E task is 55.2619 seconds. It is obvious that the individuals' FSD result in the STR C-E task is always higher than that in the RA-C task, which means the amount of the cognitive effort invested in sight translating the Chinese to English is more than that invested in reading the same Chinese text aloud.

Based on the descriptive statistics presented above, it is hypothesised that the STR tasks are much more effort-consuming than the RA tasks. In addition, a series of three paired-sample *t*-tests are conducted to corroborate this hypothesis. The Null Hypothesis is $\mu 1$ (FSD in STR) ≤ $\mu 2$ (FSD in RA) (a = 0.05). The results are shown in Table 6.1.2d. More details are displayed in Appendices (see Appendix 4, Table 6.1.2d*).

Table 6.1.2d: FSD (in Seconds) of Group A and C in RA and STR

	Group A FSD				Group C FSD	
Task	RA-C	STR C-E	RA-E	STR E-C	RA-E	STR E-G
μ	83.1721	138.4341	85.3978	135.6211	69.1186	144.5651
SD	10.8251	14.6508	9.6343	18.9142	7.2262	18.3574
p (T≤t) one tail	2.52659E-19 < 0.01		3.72845E-15 < 0.01		6.65587E-14 < 0.01	
Cohen's *d*	4.29 > 0.8		3.35 > 0.8		5.41 > 0.8	
*H0	μ1(FSD in STR) ≤ μ2(FSD in RA)(a=0.05)					

The *t*-statistics are all higher than the one-tail critical *t*-values (20.8862 > 1.6991; 14.5418 > 1.6991; 18.4816 > 1.7291) (See Appendix 4, Table 6.1.2d*). The corresponding *p*-values are all well below 0.01. Therefore, this study rejects the H_0 with a high confidence level and argues that both groups' FSD results are significantly different in the STR and the RA tasks. Furthermore, the three *d* values (Cohen's *d*) that convey the size of an effect relative to the variability in the samples are calculated respectively. They (Cohen's $d = 4.2903, 3.3461, 5.4083$) all greatly exceed the reference point of a *large* effect size (Cohen's $d = 0.8$). As a result, it is found based on the FSD results that the subjects in the two groups devoted less cognitive effort in the RA tasks, whereas they put more effort into completing the STR tasks.

In summary, in this intra-group comparison of the RA and the STR process, the amount of Group A's and Group C's cognitive effort was displayed and contrasted within the test groups with the help of the indicator FSD. Three sets of comparisons showed that both the Chinese and the German subjects in this experiment have devoted more cognitive effort to the STR tasks than to the RA tasks. This finding is in line with the conclusion obtained from the pupil diameter, and it is checked by the next indicator: total fixation duration.

6.1.2.4 Total Fixation Duration (TFD)

Total fixation duration is a common indicator of cognitive effort in eye-tracking based studies. The longer the total fixation duration, the greater the cognitive effort (Saldanha and O'Brien 2014: 144). A subject's total fixation duration in a task is the overall duration of all the fixations cast on the reading material/ST.

First of all, Group C's TFD is analysed. Group C's average total fixation duration is 58.973 seconds ($n = 20$, $SD = 5.9451$) in the RA-E task and 127.302 seconds ($n = 20$, $SD = 17.7454$) in the STR E-G task. The individuals' TFD in the two tasks are shown in Figure 6.1m. The triangles represent individuals' TFD in RA-E and the squares represent the individuals' TFD in STR E-G.

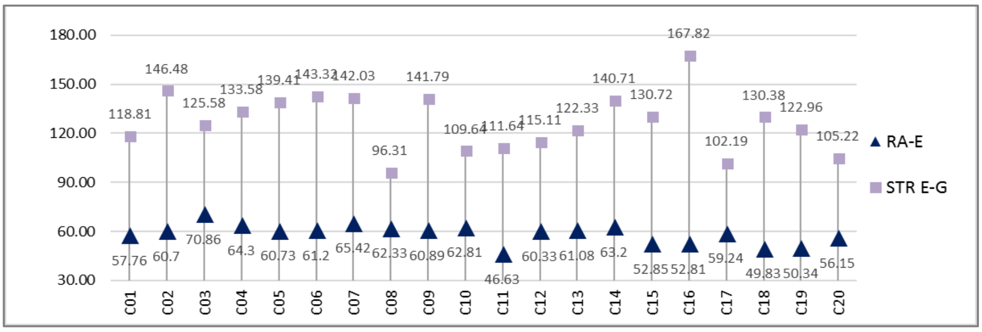


Figure 6.1m: TFD (in Seconds) of Group C in RA-E and STR E-G

Some quantitative differences are shown in the sample: every individual in Group C has longer total fixation duration in the STR task. The difference between one German subject's total fixation duration in the RA-E task and the STR E-G task is 68.3285 seconds on average. Thus, it is assumed the amount of cognitive effort invested in sight translating the English passage into German is more than that invested in reading the same piece of English passage aloud.

Following the initial finding, Group A's TFD is investigated. Group A's average total fixation duration is 76.422 seconds (n = 30, *SD* = 9.8170) in the RA-E task and 118.3427 seconds (n = 30, *SD* = 17.9661) in the STR E-C task. The individuals' total fixation duration results are presented in Figure 6.1n. The triangles represent the individuals' TFD in RA-E and the squares represent the individuals' TFD in STR E-C.

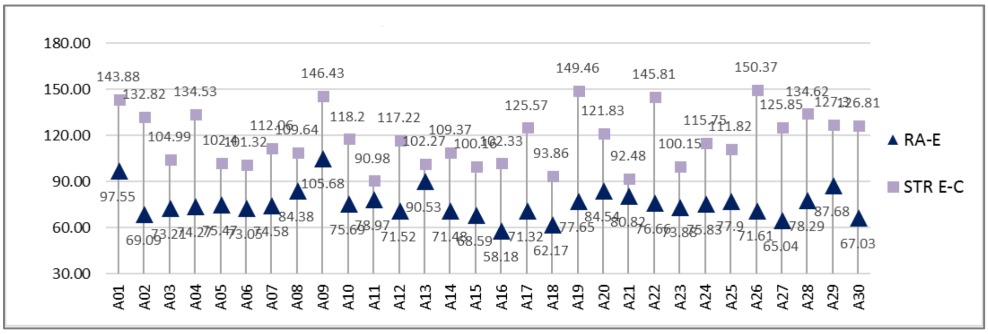


Figure 6.1n: TFD (in Seconds) of Group A in RA-E and STR E-C

The chart shows that every subject in Group A has a longer total fixation duration in the STR E-C task than in the RA-E task. On average, the difference between the total fixation duration in the RA-E task and the STR E-C task is 41.9207 seconds. As a result, the amount of attention and cognitive effort invested in sight translating English into Chinese is more than that invested in reading aloud the same English text.

Subsequently, Group A's average total fixation duration in another two tasks are compared. Their average TFD is 74.2743 seconds (n = 30, *SD* = 8.8318) in the RA-C task and 121.7827 seconds (n = 30, *SD* = 15.2596) in the STR C-E task. The individuals' total fixation durations in the RA-C task and the STR C-E task are presented in Figure 6.1o. The triangles represent the individuals' TFD in RA-C, and the squares represent the individuals' TFD in STR C-E.

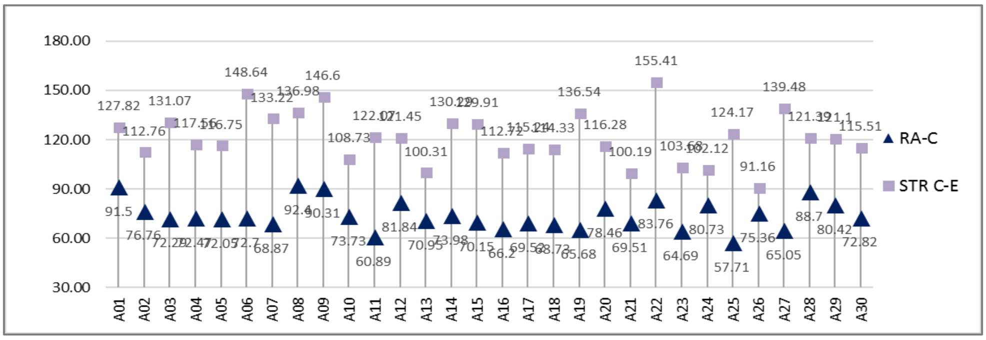


Figure 6.1o: TFD (in Seconds) of Group A in RA-C and STR C-E

On average, the difference between one Chinese subject's total fixation duration in the RA-C task and the STR C-E task is 47.5083 seconds. Although the differences vary, it is evident that the individuals' total fixation duration in the STR C-E task is always higher than that in the RA-C task, which means the amount of attention and cognitive effort invested in sight translating the Chinese passage into English is more than that invested in reading the same Chinese passage aloud.

Based on the descriptive statistics presented above, it is hypothesised that the STR tasks are much more effort-consuming than the RA tasks. Subsequently, the present study carries out three *t*-tests, in the hope of gaining support from the inferential statistics. The *t*-tests are paired and one-tail. The Null Hypothesis is $\mu 1$ (TFD in STR) $\leq$ $\mu 2$ (TFD in RA) (a = 0.05). The descriptive and inferential results are shown in Table 6.1.2e. Table 6.1.1e* in the Appendices (see Appendix 4). The tables contain more details about this *t*-test.

Table 6.1.2e: TFD (in Seconds) of Group A and C in RA and STR

	Group A TFD				Group C TFD	
Task	RA-C	STR C-E	RA-E	STR E-C	RA-E	STR E-G
μ	74.2743	121.7827	76.4220	118.3427	58.9730	127.3015
SD	8.8318	15.2596	9.8170	17.9661	5.9451	17.7454
p (T$\leq$t) one tail	1.33398E-16 < 0.01		5.76835E-14 < 0.01		4.91784E-13 < 0.01	
Cohen's *d*	3.81 > 0.8		2.90 > 0.8		5.16 > 0.8	
*H_0	$\mu 1$(TFD in STR) $\leq$ $\mu 2$(TFD in RA)(a=0.05)					

The *t*-statistics are all greater than the one-tail critical *t*-values (16.5313 > 1.6991; 13.0482 > 1.6991; 16.5318 > 1.7291) (See Appendix 4, Table 6.1.1e*). The corresponding *p*-values are all well below 0.01. Therefore, this study rejects the H_0 with a high confidence level and argues that both groups' TFD results are significantly different higher in the STR tasks. Furthermore, the three *d* values (Cohen's *d* = 3.8107, 2.8957, 5.1634) all greatly exceed the reference point of a *large* effect size (Cohen's *d* = 0.8). Based on the TFD results, it is found that the subjects devoted less cognitive effort during RA, but more cognitive effort during STR.

In summary, this section displays and compares the amount of the cognitive effort invested in the RA tasks and the STR tasks respectively, using the indicator total fixation duration. All three comparisons showed that both the Chinese and German subjects in this experiment have devoted more effort to the STR tasks than to the RA tasks. This finding is in line with the conclusion drawn from the two groups' pupil diameter and FSD results, and it is checked against the last indicator: fixation count.

6.1.2.5 Fixation Count (FC)

Since a larger number of fixations is associated with a greater investment of cognitive effort, it is employed as the last indicator in this intra-group comparison of the RA and the STR process.

The German subjects in Group C performed two tasks in the experiment: RA-E and STR E-G, which used the same English text as the reading material and the ST. As a group, their average fixation count is 256.35 (n = 20, *SD* = 22.6606) in the RA-E task and 504.6 (n = 20, *SD* = 81.6619) in the STR E-G task. The individuals' number of fixations in the two tasks are presented in Figure 6.1p. The triangles represent the individuals' FC in RA-E and the circles represent the individuals' FC in STR E-G.

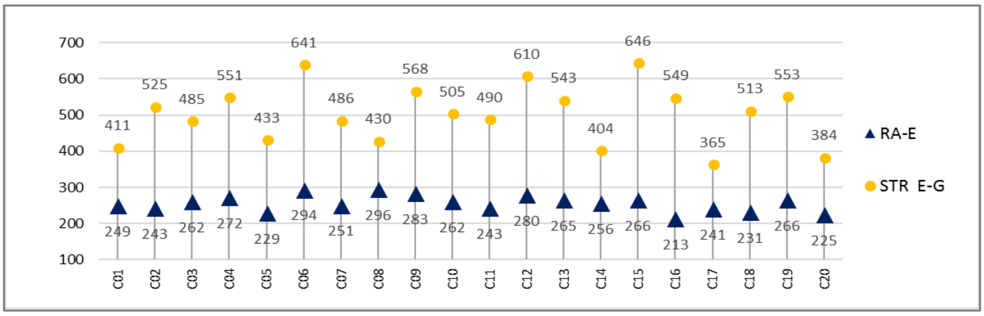


Figure 6.1p: FC (in Numbers) of Group A in RA-E and STR E-G

The remarkable difference is shown in the chart: every individual in Group C has a more substantial number of fixations in the STR task. The difference between one German subject's fixation count in the RA-E task and the STR E-G task is 248.25 on average. Thus, it is assumed that the amount of attention and cognitive effort invested in sight translating the English passage into German is more than that invested in RA.

The RA-E task and the STR E-C task presented the same English text as the reading material and the ST. On average, the Chinese subjects cast 304.2333 fixations (n = 30, *SD* = 32.3085) in the RA-E task and 491.8667 fixations (n = 30, *SD* = 86.8390) in the STR E-C task. The individuals' fixation counts in these two tasks are displayed in Figure 6.1q. The triangles represent the individuals' FC in RA-E and the circles represent the individuals' FC in STR E-C.

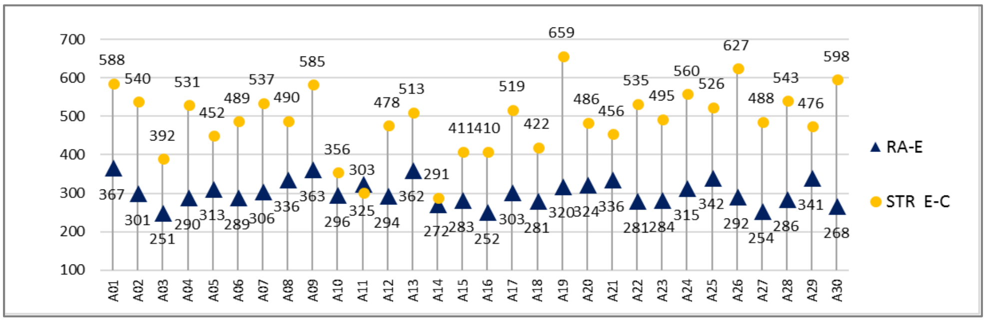


Figure 6.1q: FC (in Numbers) of Group A in RA-E and STR E-C

The chart shows that 29 subjects in Group A have a larger number of fixations in the STR E-C task. On average, the difference between one Chinese subject's fixation count in the two tasks is 187.6333. There is, however, one subject (A11) who cast more fixations on the screen during RA-E. Nevertheless, it is hypothesised that the amount of attention and cognitive effort invested in sight translating the English passage into Chinese is more than that invested in reading the same English passage aloud.

Lastly, Group A's fixation counts in the RA-C task and the STR C-E task are analysed to test the initial hypothesis. Group A's average fixation count is 283.8333 (n = 30, *SD* = 27.2524) in the RA-C task and 470.3667 (n = 30, *SD* = 70.4177) in the STR C-E task. The individuals' fixation counts in these two tasks are displayed in Figure 6.1r. The triangles represent the individuals' FC in RA-C and the circles represent the individuals' FC in STR C-E.

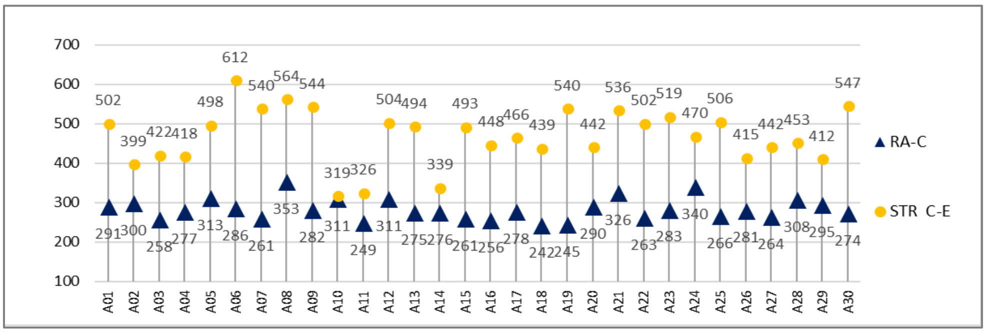


Figure 6.1r: FC (in Numbers) of Group A in RA-C and STR C-E

By comparing their fixation count in the RA-C and STR C-E tasks, which presented the same Chinese text as the reading material and the ST, it is shown that the Chinese subjects cast more fixations on sight translating the text than reading it. On average, the difference between one Chinese subject's fixation count in the two tasks is 186.5333. Although there is one subject (A10) whose number of fixations in the STR C-E task is eight more than his/her number of fixations in the RA-C task, it is still evident that the number of fixations in the STR C-E task is higher than that in the RA-C task. Hence, this data means the amount of attention and cognitive effort invested in sight translating the Chinese passage into English is larger than that invested in reading the same piece of Chinese passage aloud.

Despite a couple of exceptions, it is hypothesised that the STR tasks are much more effort-consuming than the RA tasks based on the descriptive statistics presented above. Three one-tail paired *t*-tests were conducted in order to gain support from inferential statistics. The Null Hypothesis is $\mu 1$ (FC in STR) $\leq \mu 2$ (FC in RA) ($a = 0.05$). Both descriptive and inferential data are presented in Table 6.1.2f. Table 6.1.1f* in the Appendices (see Appendix 4) and contain more details about this *t*-test.

Table 6.1.2f: FC (in Numbers) of Group A and C in RA and STR

	Group A FC				Group C FC	
Task	RA-C	STR C-E	RA-E	STR E-C	RA-E	STR E-G
μ	283.8333	470.3667	304.2333	491.8667	256.3500	504.6000
SD	27.2524	70.4177	32.3085	86.8390	22.6606	81.6619
p (T≤t) one tail	6.33026E-15 < 0.01		2.27691E-13 < 0.01		3.02274E-12 < 0.01	
Cohen's *d*	3.49 > 0.8		2.86 > 0.8		4.14 > 0.8	
*H_0	$\mu 1$(FC in STR) $\leq \mu 2$(FC in RA)($a=0.05$)					

Table 6.1.2f* (see Appendix 4) shows that the *t*-statistics are all greater than the one-tail critical *t*-values (14.2436 > 1.6991; 12.3431 > 1.6991; 14.9202 > 1.7291). The corresponding *p*-values are all well below 0.01. Therefore, this study can reject the H_0 with a high confidence level. The three *d* values

(Cohen's *d* = 3.4937, 2.8639, 4.1426), which convey the size of an effect relative to the variability in the samples, all exceed the reference point of a *large* effect size (Cohen's *d* = 0.8). Therefore, it is concluded that the subjects devoted more cognitive effort to completing the STR tasks.

In summary, this section displays and contrasts the amount of the cognitive effort invested in the RA tasks and the STR tasks respectively, using the indicator fixation count. This indicator shows a significant difference between the RA tasks and the STR tasks, no matter which experiment groups are compared. Compared to the RA tasks, the STR tasks receive a greater number of fixations "because the eyes were required not only to feed the brain with input for meaning construction, but also to supply the brain with online monitoring information about what portions of text had been satisfactorily covered by the spoken translation output and what elements remained to be dealt with" (Jakobsen and Jensen 2008: 117). It is stated explicitly by this indicator that when the material for the RA task and the STR task is the same text, all the Chinese and German subjects spent a lot more effort in completing the STR task. Furthermore, this finding is in line with the conclusion drawn from the two group's pupil diameter, TTT, FSD, and TFD results.

In conclusion, this study carries out more in-depth, intra-group comparisons within Group A and Group C in Section 6.1.2 with the aim of examining the second hypothesis using pupil diameter, TTT, FSD, TFD, and FC. In general, despite some individual exceptions, all of the four indicators are indicative of fundamental behavioural differences between subjects' RA and STR processes. Therefore, after carrying out the intra-group comparisons of two groups' eye-tracking data, three hierarchies of their investment of cognitive effort are formed: Group A STR C-E > Group A RA-C, Group A STR E-C > Group A RA-E, Group C STR E-G > Group C RA-E. Both the Chinese subjects and the German subjects invested more cognitive effort in the process of STR than in the process of RA. Hence, the analyses support the hypothesis that subjects invest different amounts of cognitive effort during

STR and RA.

6.2 Examining Mean Fixation Duration (MFD)

There are some studies that used both Fixation Count (FC) and Mean Fixation Duration (MFD) as indicators of translators' cognitive workload. Two of the studies, carried out by Sharmin *et al*. (2008) and Jakobsen and Jensen (2008) respectively, displayed some disadvantages of applying MFD.

Sharmin *et al*. (2008: 48) compared the distribution of the visual effort on ST and the TT. By investigating how time pressure and text complexity affect translators' visual attention, they found that as the text complexity increases, more fixations were required. What they disregarded in their discussion was how MFD varied with ST complexity. However, a recalculation of their data showed that, in fact, the least difficult ST in Sharmin *et al*.'s (2008) translation experiment yielded the longest MFD. On the other hand, Jakobsen and Jensen's (2008: 115) paired-sample *t*-test for the fixation count showed significant differences across four tasks, but the paired-sample *t*-test for the MFD does not. Their one-way ANOVA analysis also showed that the translation students' fixation count was significantly higher than the professional translators' fixation count. The difference between the two test groups' MFD results, however, was not considered significant.

Holmqvist *et al*.'s (2011: 416) observation may reveal the reason behind this: MFD is easily biased upwards due to a small number of long fixations. Based on their findings of the use of the MFD as an indicator, the present study has chosen to disregard MFD in previous sections. Instead, MFD is put under close inspection in Section 6.2 through both cross-group and intra-group comparisons.

Mean fixation duration is said to correlate positively with the viewing distance: the further the eyes are from the stimulus, the longer the MFD (Morrison 1983: 37). Morrison (1983: 37) argued that this was because the stimulus is less discriminable with an increased viewing distance. Therefore, in order not to let the viewing distance affect the subjects' MFD results, the subjects were all required to sit at approximately the same distance (60-65 cm) from the screen and maintain that distance during the experiments.

6.2.1 A Cross-group Comparison of MFD in RA-E

In the preliminary cross-group comparison of three student groups' eye-tracking data recorded during RA-E, a hierarchy of their investment of cognitive effort was formed earlier in Section 6.1.1: Group A > Group C ≥ Group B. This hierarchy is used as the foundation of the discussion in Section 6.2.1.

The eye-tracking recordings, which have passed the data quality assessment, are further analysed to examine the Chinese, German, and British subjects' MFD during RA-E, a task all three groups have taken part in. During RA-E, the mean MFD of Group B is 0.2354 seconds (n = 35, *SD* = 0.0298) when the subjects read aloud in their L1. It is higher when the subjects in Group A read aloud in their L2. Their mean MFD is 0.2512 seconds (n = 30, *SD* = 0.0167). The mean MFD of Group C is the lowest. It is 0.2327 seconds (n = 20, *SD* = 0.0243).

As a result, the average MFD results do not provide enough evidence to argue whether the native speakers in Group B devoted the least amount of cognitive effort to the RA task. This conclusion, nevertheless, is based solely on the descriptive statistics. Potentially, some individuals' MFD scores could have distorted the average results either positively or negatively. Therefore, a series of three independent two-sample *t*-tests assuming equal variances are conducted to corroborate whether there are significant differences among the three groups. Table 6.2a displays the results.

Table 6.2a: Two-sample T-tests Assuming Equal Variances [MFD of Group A, B, and C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

MFD in RA-E	*Group A*	*Group B*	*Group B*	*Group C*	*Group A*	*Group C*
Mean	0.251185	0.235395	0.235395	0.232711	0.251185	0.232711
Variance	0.00028	0.000889	0.000889	0.000589	0.00028	0.000589
Observations	30	35	35	20	30	20
Pooled Variance	0.000609		0.000781		0.000402	
Hypothesised Mean Difference	0		0		0	
df	63		53		48	
t Stat	2.57255		0.342619		3.191444	
P(T<=t) one-tail	0.006232		0.36662		0.001249	
t Critical one-tail	1.669402		1.674116		1.677224	
P(T<=t) two-tail	0.012464		0.73324		0.002497	
t Critical two-tail	1.998341		2.005746		2.010635	

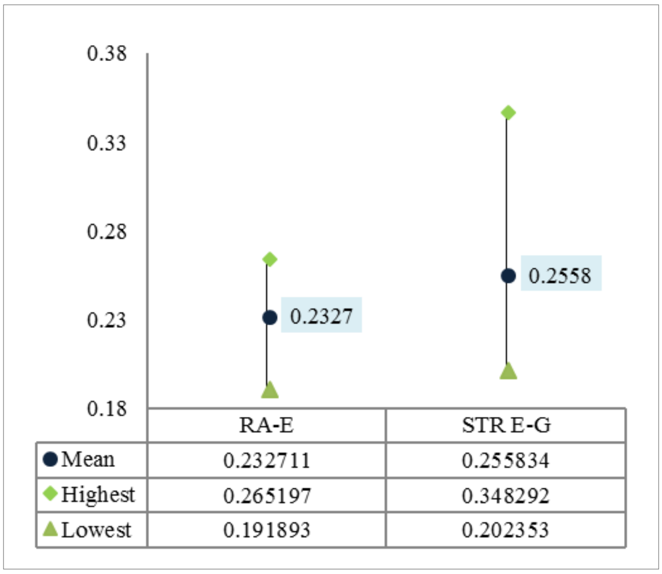
Based on the low *p*-values ($p < 0.01$), the first and the third *t*-tests show that Group A have a significantly longer MFD compared to Group B and Group C. The corresponding effect size measures are calculated (Cohen's $d = (M_2 - M_1) / SD_{\text{pooled}}$). For the first *t*-test, the standardised effect size (Cohen's $d = 0.6532$) is between the reference point of a *medium* (0.5) and a *large* effect (0.8). For the third *t*-test, the standardised effect size (Cohen's $d = 0.8865$) exceeds the reference point of a *large* effect size. The second *t*-test conducted for the MFD of Group B and Group C, on the other hand, yields a relatively high *p*-value ($p > 0.1$). The present study does not have a high confidence level to argue that there is any significant difference between Group C's and Group B's MFD. The Cohen's *d* result of the second *t*-test is 0.0987. As mentioned before, if the two dataset's means do not differ by 0.2 standard deviations or more, the difference is trivial, even if the *t*-test shows that they are statistically significant. A Cohen's *d* of 0.0987, in this case, suggests that the effect size is indeed trivial.

This means that if the comparison is only made between the group of Chinese subjects and either of the other two groups, MFD can be a useful indicator. When it comes to the comparison between the Group B and Group C, the MFD fails to indicate any statistically significant difference. Since the experimental cross-group analysis left us with some doubts about MFD, it is tested again in Section 6.2.2, where this research further examines this indicator in a comprehensive intra-group comparison between the RA and the STR tasks.

6.2.2 An Intra-group Comparison of MFD in RA and STR

In Section 6.1.2, the present study carried out in-depth, intra-group comparisons within Group A and Group C. Based on pupil diameter, FSD, TFD, and FC, three hierarchies of the investment of cognitive effort came to light: Group A STR C-E > Group A RA-C; Group A STR E-C > Group A RA-E; Group C STR E-G > Group C RA-E. In other words, both the Chinese subjects and German subjects invested more cognitive effort during STR. This hypothesis and the three hierarchies are employed as the foundation of the analysis in this section.

The intra-group comparison takes place in three steps: studying the descriptive statistics, double-checking the individual data, and considering the inferential statistics. First of all, the average results of the two tasks carried out by Group C are analysed. Figure 6.2a is a general presentation of Group C's MFD for both tasks. The circles represent the mean MFD results, the rhombuses represent the highest MFD results, and the triangles represent the lowest MFD results in the two tasks.

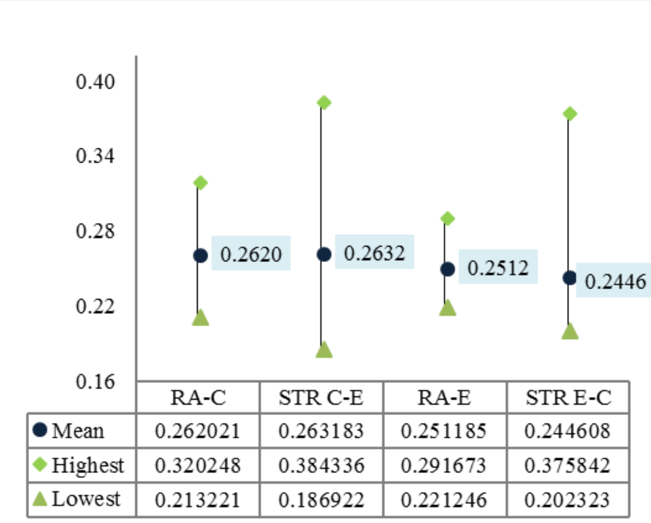


	RA-E	STR E-G
● Mean	0.232711	0.255834
◆ Highest	0.265197	0.348292
▲ Lowest	0.191893	0.202353

Figure 6.2a: MFD (in Seconds) of Group C in RA and STR

The average MFD results do not differ too much between these two tasks. The average MFD is 0.2327 seconds in the RA-E task (n = 20, *SD* = 0.0243) and 0.2558 seconds in the STR E-G task (n = 20, *SD* = 0.042). It seems that the differences between the average MFD scores show little contrast.

Subsequently, the present study investigates the Chinese subjects' MFD. Figure 6.2b displays Group A's average MFD results in the four tasks (RA-C, STR C-E, RA-E, STR E-C). The circles represent the mean MFD results, the rhombuses represent the highest MFD results, and the triangles represent the lowest MFD results in the four tasks.



	RA-C	STR C-E	RA-E	STR E-C
● Mean	0.262021	0.263183	0.251185	0.244608
◆ Highest	0.320248	0.384336	0.291673	0.375842
▲ Lowest	0.213221	0.186922	0.221246	0.202323

Figure 6.2b: MFD (in Seconds) of Group A in RA and STR

Again, the MFD results for the Chinese subjects do not vary much across the four tasks. The average MFD is 0.2512 seconds in the RA-E task (n = 30, *SD* = 0.0167), 0.2446 seconds in the STR E-C task (n = 30, *SD* = 0.0388), 0.2620 seconds in the RA-C task (n = 30, *SD* = 0.0262), and 0.2632 seconds in the STR C-E task (n = 30, *SD* = 0.0474). In fact, Group A's MFD in the RA-E task is slightly higher than that in the STR E-C task, but their MFD in the RA-C task is slightly lower than that in the STR C-E task.

By looking at the average MFD results of two groups, one could suggest the following:

a. Since Group C’s average MFD during RA-E is shorter than that during STR E-G, a longer MFD represents a larger amount of cognitive effort invested in the task.

b. Since Group A’s average MFD during RA-E is longer than that during STR E-C, a shorter MFD represents a larger amount of cognitive effort invested in the task.

c. Since Group A’s average MFD during RA-C is shorter than that during STR C-E, a longer MFD represents a larger amount of cognitive effort invested in the task.

These three conclusions are contradictory, making it unadvisable for the researcher to draw a conclusion from the average MFD. Since the differences between the average MFD for each task display little contrast, individual results are compared one by one. In other words, instead of looking at a group’s average MFD scores in the RA and the STR tasks, every subject’s MFD score in a RA task is compared with their own MFD score in the corresponding STR task.

Group C’s results are compared first. Instead of looking at a group’s average MFD, every subject’s MFD in the RA-E task is compared with the MFD in the STR E-G task. Figure 6.2c displays the results in both tasks. The bars represent the subjects’ average MFD in the RA-E task, while the squares represent their average MFD in the STR E-G task.

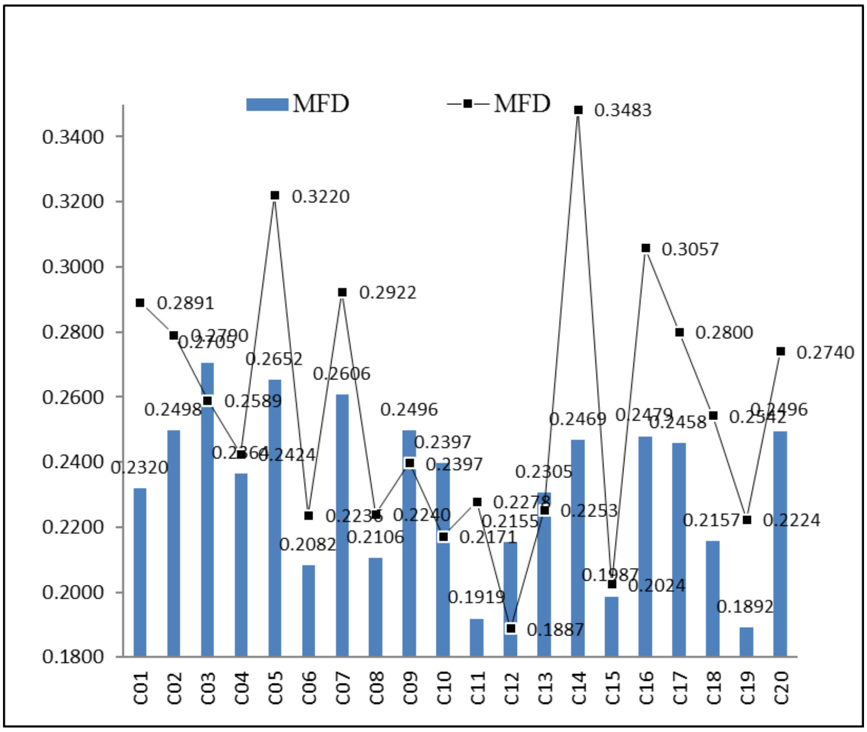


Figure 6.2c: Individual MFD (in Seconds) of Group C in RA-E and STR E-G

Figure 6.2d shows the Chinese subjects’ MFD results in all four tasks. The Chinese subjects’ average MFD results in the RA-C task are represented by the bars and their average MFD results in the other tasks are represented by the different black shapes: circles for the MFD results in the STR C-E task, squares for the MFD results in the STR E-C task, and triangles for the MFD results in the RA-E task.

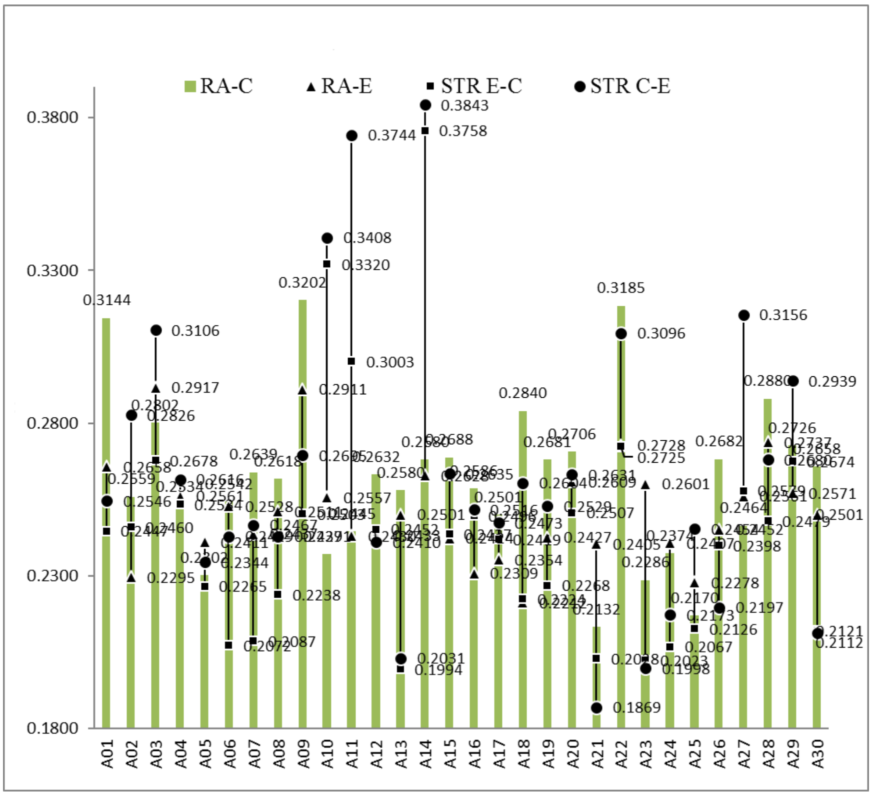


Figure 6.2d: Individual MFD (in Seconds) of Group A in RA and STR

The interpretation of the Chinese subjects' and German subjects' MFD results presented above lead to the following conclusions:

1. Group C subjects have a slightly higher mean MFD in the STR E-G task than in the RA-E task when reading their L2 text. Out of the 20 German subjects in Group C, 15 have longer fixations during STR E-G than during RA-E. Since the STR E-G task is more effort-consuming than the RA-E task, one could hypothesise that these German speakers had shorter fixations in the task that demanded less cognitive processing.
2. Group A subjects have a slightly lower mean MFD in the STR E-C task than in the RA-E task when reading their L2 text. 63.33% of the individuals have a lower MFD during STR E-C than during RA-E. Therefore, the mean MFD of Group A represents the majority of the group. Since the STR E-C task is more effort-consuming than the RA-E task, one could hypothesise that these Chinese speakers had longer fixations in the task that demanded less cognitive processing.

3. Group A subjects have a slightly higher mean MFD in the STR C-E task than in the RA-C task when reading their L1 text. However, only 33.33 % of the individuals have a higher MFD during STR C-E than during RA-C. Therefore, the mean MFD of Group A does not represent the majority of the group. As a result, most of the Chinese subjects have longer fixations in the RA-C task than in the STR C-E task. Since the STR C-E task is more effort-consuming than the RA-C task, one could hypothesise that these Chinese speakers had longer fixations in the task that demanded less cognitive processing.

In summary, the analysis of the German subjects' MFD suggests that a shorter MFD is associated with a lower level of cognitive effort investment, which is in agreement with other researchers in previous studies (e.g., Inhoff and Rayner 1986; Ehrlich and Rayner 1981; Rayner and Pollatsek 1989). However, the analysis of Chinese subjects' MFD shows that a higher MFD score is associated with a lower level of cognitive effort investment, which contradicts the abovementioned researchers' point of view that having longer fixations indicates a higher level of perceived task difficulty and a more substantial amount of cognitive effort devoted to the task.

The fact that the descriptive statistics and the individual results cannot provide a consistent explanation for MFD leads the discussion to the final stage, which involves analysing the inferential statistics of the two test groups. A paired two-sample *t*-test and a single factor ANOVA test are conducted with Group C's and Group A's MFD results respectively. The results are presented in Table 6.2b and Table 6.2c.

Table 6.2b: Paired Two-sample T-test for Means [MFD of Group C in RA-E and STR E-G]

T-Test: Paired Two Sample for Means

	STR E-G	*RA-E*
Mean	0.255834	0.232711
Variance	0.001763	0.000589
Observations	20	20
Pearson Correlation	0.667378	
Hypothesised Mean Difference	0	
df	19	
t Stat	3.283447	
P(T<=t) one-tail	0.001954	
t Critical one-tail	1.729133	
P(T<=t) two-tail	0.003908	
t Critical two-tail	2.093024	

Table 6.2c: A Single Factor ANOVA Test: MFD of Group A in RA and STR

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.007157	3	0.002386	2.025787	0.114126	2.682809
Within Groups	0.136599	116	0.001178			
Total	0.143756	119				

The paired two-sample for means *t*-test suggests that the MFD of the German subjects in Group C is statistically higher in the more effort-consuming task: STR E-G. The *d* value (Cohen's $d = 0.6742$) that conveys the size of an effect relative to the variability in the samples is a *medium* effect size. On the other hand, the high *p*-value ($p > 0.1$) in the ANOVA test suggests that the confidence level of rejecting the null hypothesis is not high, meaning that there is not a significant difference between Group A's MFD across the four different tasks. Since it is not affected by the independent variable, researchers are unable to tell which task of these four consumed the most or the least amount of cognitive effort by looking at the Chinese subjects' MFD results.

Hvelplund (2014: 204) considered the following question: "Can we, for instance, be certain that longer fixations actually reflect more processing intensity?". In Section 6.2.1, the preliminary analysis of the three test groups'

MFD in RA-E questions whether MFD is a valid and reliable indicator of the cognitive effort invested into a visual task. After starting to question the validity and reliability of MFD, Section 6.2.2 goes into a detailed examination of the use of MFD in eye-tracking research through intra-group comparisons between the RA and the STR tasks. However, the conclusions from the three steps of analysis are not consistent.

Indeed, fixation is shown to be a useful indicator of mental activities related to text processing (e.g., Rayner 1978) and it is relatively well accepted that total fixation duration is taken as an indicator of processing time (Shebilske and Fisher 1983: 173). Mean fixation duration, however, does not equal total fixation duration. Carpenter and Just (1983: 277) stated in their work that "the time that a reader spends on a word reflects processes initiated by that word". Their argument provides support for using total fixation duration, rather than the mean fixation duration. When the overall processing time of a given word or phrase is set, the reader could reach the required threshold by either casting a single long fixation or a couple of shorter fixations. One single fixation does not allow sufficient acuity on a long word (Carpenter and Just 1983: 287). Therefore, the total fixation duration on a word increases with the word length, not because any single fixation is extended, but because the number of fixations increases. Therefore, these two measures employed to estimate the investment of mental resources deserve more extensive distinction and justification. Having a larger number of fixations is interpreted as the task being more cognitively demanding, while having a higher number of fixations within a specific period results in having shorter fixations. Therefore, having longer fixations cannot be a perfect indication of investing more cognitive effort.

The MFD result in this study does not show validity, because this dependent variable does not reflect what the researcher thought it would measure. It does not have reliability either, as this chosen measure fails to consistently support the tentative conclusions. Instead, MFD seems to be rather *idiosyncratic*, which means individual subjects have their own basic setting for the value, and thereby the difference between subjects is much more substantial than the variation within the subjects when the experimental conditions were changed (Holmqvist *et al*. 2011: 83). If researchers rely on MFD alone, their research may be biased. Nevertheless, it is not my intention to argue that MFD should be excluded as a measure from all eye-tracking based studies. This study does not intend to question the results of eye-tracking based studies that have applied MFD successfully. The results they have obtained are indeed very important and intriguing. Therefore, this research is suggesting that researchers dismiss MFD, but to execute caution when applying it as a measurement of cognitive effort.

6.3 Exploring Two New Indicators

Although it is generally believed that "the pattern of fixations is sensitive to the types of mental processes occurring during reading" (Hogaboam 1983: 331), researchers should not limit their choice to traditional fixation-based indicators. This section is dedicated to exploring two potential indicators and their relationship with the mental processes occurring during reading-related tasks. Fixation Time on Source text as a percentage of total task time (FTS) and Saccade duration as a Percentage of pure Gaze activities (SPG) were used as criteria for the eye-movement quality assessment in Chapter 5. In this section, the present study intends to reveal their potential to become new indicators of cognitive effort.

6.3.1 Fixation Time on Source Text as a Percentage of Total Task Time (FTS)

Thirty-five years ago, the assumption that a longer *fixation durations per second* indicates that more mental effort was invested to encode the visual information was not supported by empirical evidence (Inhoff 1983: 191). In a more recent study, Jakobsen and Jensen (2008) attempted to re-examine this measurement. They rebranded the measurement as *average gaze times in percent of total task time* and employed it as a measurement for measuring the cognitive load spent in reading and translation tasks. By looking at the average gaze/task time value of the two reading tasks in Jakobsen and Jensen's (2008) research, one could see that both professional translators and translation students had a higher gaze/task time value in the *reading for general comprehension* task than in the *reading for translation* task. In other words, subjects have more condensed fixations on the reading material when the task is less demanding. Average gaze times in percent of total task time is similar to FTS. The only difference is that the fixation time on source text excludes marginal fixations that did not fall on the ST. Since these two measurements are so similar, is it possible that Jakobsen and Jensen's (2008) hypothesis coincides with the finding of the FTS in the present research?

Another relevant measure has been addressed by researchers relatively recently. *Fixation Rate*, which refers to the number of fixations per unit time, is calculated as the number of fixations per second (s^{-1}) (Holmqvist *et al*. 2011: 416). It was "explicitly used as measure in its own right" (ibid.) and was found to correlate negatively with mental demands (Nakayama *et al*. 2002): the more demanding the task, the lower the fixation rate. Since the total fixation duration correlates positively with the fixation count, it is reasonable to assume that the total fixation duration per unit time also correlates negatively with the demands of mental processing.

In the present study, Fixation Time on Source text as a percentage of total task time (FTS) refers to the percentage of the task time that is spent entirely on fixations. FTS = [(total fixation duration on the source text ÷ total task time) × 100]. First, the three groups' FTS results in the RA-E task are shown in Figure 6.3a.

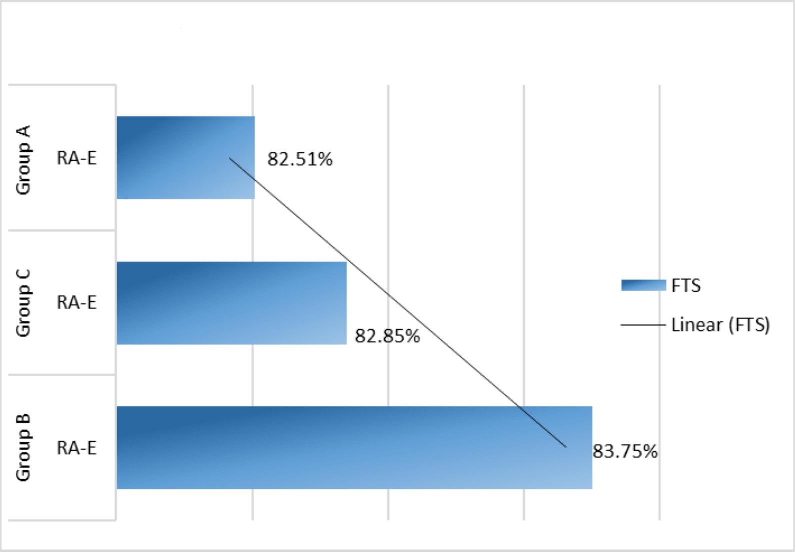


Figure 6.3a: Average FTS (in Percentages) of Group A, B, and C in RA-E

In the preliminary cross-group comparison, a hierarchy of their investment of cognitive effort during RA-E was formed in Section 6.2.1: Group A > Group C ≥ Group B. Since the group that invested the least amount of cognitive effort in the RA-E task has the highest FTS score, the present researcher hypothesises that the higher the FTS, the lower the cognitive effort.

Admittedly, looking at the groups' average FTS is not sufficient to account for the relationship between FTS and the cognitive demands of a task. Hence, to test this initial hypothesis and to determine whether FTS could be a potential indicator of the investment of cognitive effort in the RA and the STR tasks, Group A's and Group C's FTS results shall go through a much more thorough investigation than simply looking at the two groups' average results. In short, Group A's FTS is 82.04% in the RA-C task (n = 30, *SD* = 0.0469), 82.51% in the RA-E task (n = 30, *SD* = 0.0408), 74.87% in the STR C-E task (n = 30, *SD* = 0.0697), and 75.88% in the STR E-C task (n = 30, *SD* = 0.0614). Thereafter, Group C's FTS scores are compared between the two tasks. In summary, Group C's FTS is 82.85% in the RA-E task (n = 20, *SD* = 0.0449) and 76.97% in the STR E-G task (n = 20, *SD* = 0.0756). Again, the average FTS suggests that the subjects tend to have a lower FTS in the STR tasks. Since the STR tasks are more cognitively demanding and more effort-consuming than the corresponding RA tasks, the initial hypothesis is confirmed.

Figure 6.3b displays Group C's FTS results in the two tasks. The bars represent the average FTS in the RA-E task and the black squares represent the average FTS in the STR E-G task.

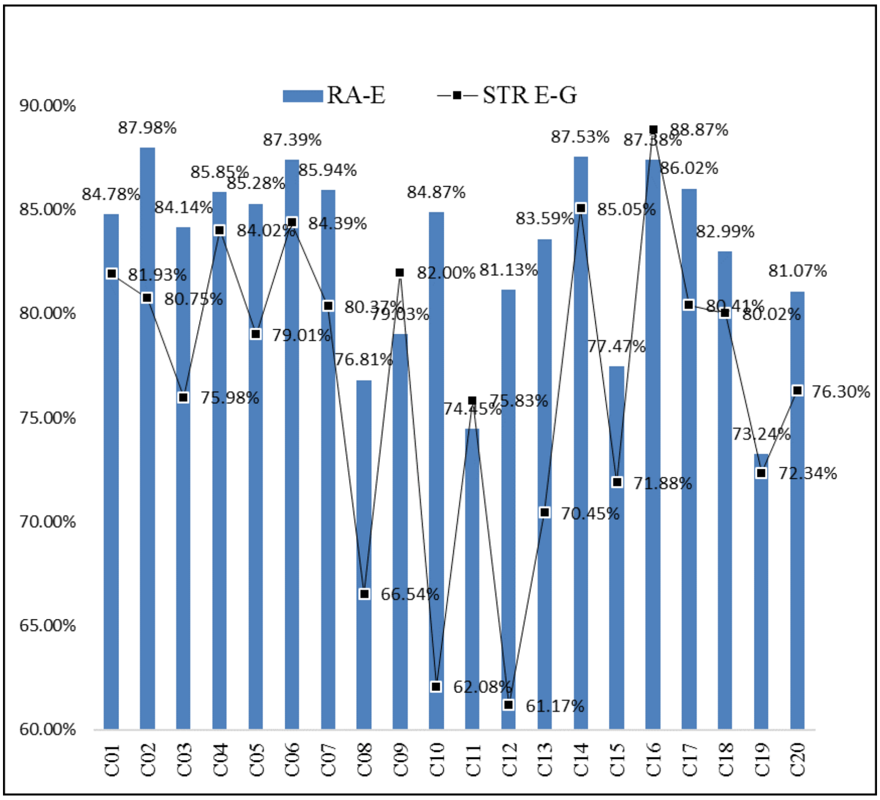


Figure 6.3b: Individual FTS (in Percentages) of Group C in RA and STR

The same tendency is found in Group C's FTS results: the majority (85%) of the subjects in Group C has a higher FTS in the RA-E task. Assuming that these subjects have a significantly higher FTS during RA-E than during STR E-G, a one-tailed and paired *t*-test is conducted to corroborate this hypothesis. The corresponding Null Hypothesis is $\mu 1$ (FTS in RA-E) $\leq \mu 2$ (FTS in STR E-G) (a = 0.05). Table 6.3a shows the results.

Table 6.3a: Paired Two-sample T-test for Means [FTS of Group C in RA-E and STR E-G]

T-Test: Paired Two Sample for Means

Group C	*Variable 1 (FTS in RA-E)*	*Variable 2 (FTS in STR E-G)*
Mean	0.82847731	0.769698834
Variance	0.002015102	0.005710927
Observations	20	20
Pearson Correlation	0.496013036	
Hypothesised Mean Difference	0	
df	19	
t Stat	3.98065203	
P(T<=t) one-tail	0.000400407	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	0.000800815	
t Critical two-tail	2.093024054	

Table 6.3a shows that the t-statistic (3.980652) is higher than the one-tail critical t-value (1.729133). The corresponding p-value is approximately 0.0004, which is significantly lower than 0.01. There is an exceptionally low probability that the difference between their FTS in the RA and the STR tasks occured accidentally. The low p-value ($p < 0.01$) coming from the paired two-sample for the means t-test gives the present research an extremely high confidence level to reject the H_0. Also, the standardised effect size (Cohen's $d = 0.9457$) exceeds the reference point of a *large* effect (0.8). Therefore, the subjects in Group C have a statistically higher FTS in the RA task than in the STR task.

Figure 6.3c shows Group A's FTS results in all four tasks. The Chinese subjects' average FTS results in the RA-C task are represented by the bars in the chart. The FTS results in the other tasks are represented by the different black shapes. The circles represent the FTS in the STR C-E task, the squares represent the FTS in the STR E-C task, and the triangles represent the FTS in the RA-E task.

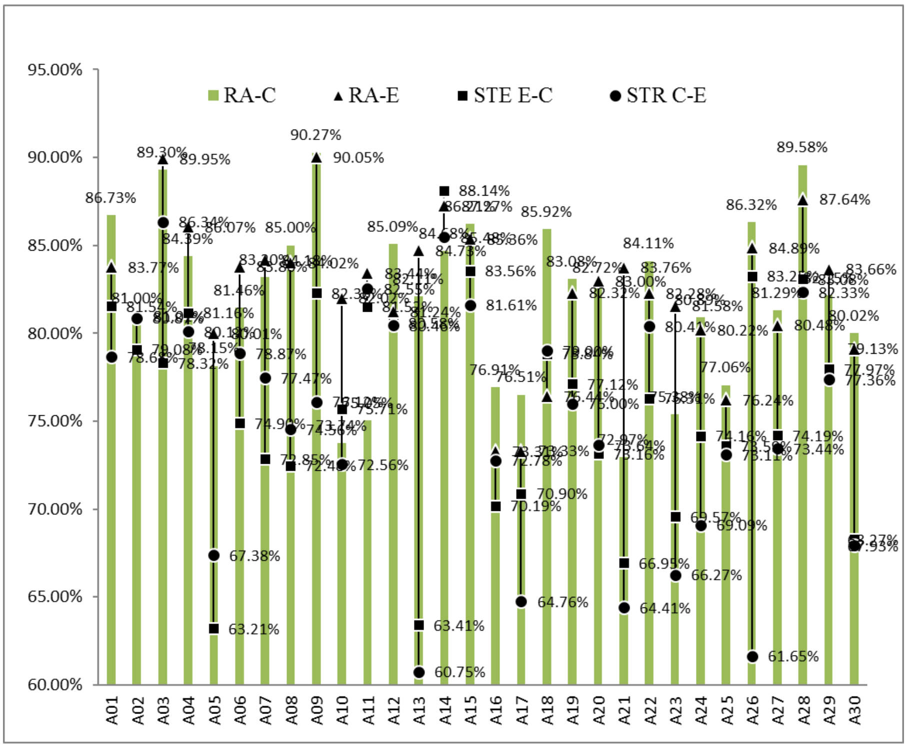


Figure 6.3c: Individual FTS (in Percentages) of Group A in RA and STR

In summary, 93.33% of the group has a higher FTS in the RA-C task than in the STR C-E task; 90% of the group has a higher FTS in the RA-E task than in the STR E-C task. Assuming that the subjects in Group A have a significantly higher FTS during RA-C than during STR C-E and a significantly higher FTS during RA-E than during STR E-C, two one-tailed and paired *t*-tests are carried out. The corresponding Null Hypotheses are $\mu 1$ (FTS in RA-C) $\leq \mu 2$ (FTS in STR C-E) (a = 0.05) and $\mu 1$ (FTS in RA-C) $\leq$ $\mu 2$ (FTS in STR C-E) (a = 0.05). Table 6.3b and Table 6.3c display the results.

Table 6.3b: Paired Two-sample T-test for Means [FTS of Group A in RA-C and STR C-E]

T-Test: Paired Two Sample for Means

Group A	*Variable 1 (FTS in RA-C)*	*Variable 2 (FTS in STR C-E)*
Mean	0.820433902	0.748666327
Variance	0.002194887	0.004851959
Observations	30	30
Pearson Correlation	0.487353579	
Hypothesised Mean Difference	0	
df	29	
t Stat	6.322027601	
P(T<=t) one-tail	3.30484E-07	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	6.60968E-07	
t Critical two-tail	2.045229642	

Table 6.3c: Paired Two-sample T-test for Means [FTS of Group A in RA-E and STR E-C]

T-Test: Paired Two Sample for Means

Group A	*Variable 1 (FTS in RA-E)*	*Variable 2 (FTS in STR E-C)*
Mean	0.825060158	0.758768614
Variance	0.00166212	0.003767349
Observations	30	30
Pearson Correlation	0.464495177	
Hypothesised Mean Difference	0	
df	29	
t Stat	6.516308605	
P(T<=t) one-tail	1.95227E-07	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	3.90455E-07	
t Critical two-tail	2.045229642	

The *t*-statistics are much higher than the critical *t*-values in both of the tables. Meanwhile, the corresponding *p*-values are significantly lower than 0.01. Hence, the present research has an exceptionally high confidence level to reject the H_0. The *p*-values ($p < 0.01$) coming from the two-paired, two-sample *t*-tests confirm that Group A has a statistically higher FTS in the RA tasks than in the STR tasks. Furthermore, the standardised effect size measures of the two *t*-tests are calculated. In the first case, the *d* value that conveys the size of an effect relative to the variability in the samples is 1.2091; in the second case, the standardised effect size Cohen's *d* is 1.2723. Both results suggest that the effect sizes are *large*.

In conclusion, based on the hierarchy of the cognitive effort investment formed in Section 6.1, the relationship between FTS and cognitive processing is examined in this section. Both the descriptive and the inferential statistics show that a higher FTS indicates a smaller amount of cognitive effort devoted to the visual task. One could argue that the subjects might have looked away from the screen occasionally. This, however, is rather unlikely because an STR task requires the subjects to focus on the screen quite intensively within a short amount of time. Based on the analysis above, the present study suggests that having to translate the text into another language reduces FTS value because it increases the active eye movements between fixations: saccades. Therefore, the discussion focuses on SPG in Section 6.3.2, hopes to support this hypothesis by measuring the percentage of the saccades directly.

6.3.2 Saccade Duration as a Percentage of Pure Gaze Activities (SPG)

The control of eye movements determines what visual information is available to the reader, in which order the critical information is obtained, and for how long the information is available (McConkie 1983: 67). It is, however, not restricted to the control of the fixation-length and the fixation-location. Rather, it also involves the control and management of the saccades. Hence, the present study wishes to explore a saccade-related measurement as an independent indicator.

Saccade duration as a Percentage of pure Gaze activities (SPG) defines how many percentage points of the duration of pure gaze activities are entirely attributed to saccades. SPG = [(saccade duration ÷ the sum of fixation and saccade duration) × 100]. As stated in Chapter 5, the normal distribution of one's total saccade duration as a percentage of the sum of the overall saccade and fixation duration is between 5% and 15%.

It is hypothesised that this saccade-related measurement could be used as an independent indicator. Firstly, individual subjects' FTS and SPG results in the RA-E task are analysed. In fact, the Pearson correlation coefficient (*r*) is -0.6783 between Group A's FTS and SPG; -0.7658 between Group B's FTS and SPG; -0.8109 between Group C's FTS and SPG. These negative *r* values suggest that there is a negative correlation between FTS and SPG. Since both the descriptive and the inferential statistics in the previous section show that a higher FTS indicates a lower amount of cognitive effort devoted to the visual task, the present research hypothesises that the higher the SPG, the larger the investment of cognitive effort.

Group A's and Group C's SPG results undergo a thorough investigation to test this hypothesis. Again, the investigation is based on three hierarchies of cognitive effort formed in Section 6.1.2 (Group A STR C-E > Group A RA-C, Group A STR E-C > Group A RA-E, Group C STR E-G > Group C RA-E).

The mean SPG results show a clear contrast between RA and STR. On average, the Chinese subjects have a lower SPG in the RA-C ($\mu = 9.09\%$, n = 30, *SD* = 0.0195) and the RA-E tasks ($\mu = 9.35\%$, n = 30, *SD* = 0.0116) and a higher SPG in the STR E-C ($\mu = 12.18\%$, n = 30, *SD* = 0.0177) and the STR C-E tasks ($\mu = 11.53\%$, n = 30, *SD* = 0.0254). Meanwhile, the German subjects' mean SPG is 10.50% (n = 20, *SD* = 0.0221) in the RA-E task and 11.69% (n = 20, *SD* = 0.0195) STR E-G task. The results show that they have a higher percentage of saccade during STR E-G. Since the STR tasks are more cognitively demanding and more effort-consuming than the corresponding RA tasks, the initial hypothesis is confirmed.

When it comes to the individual data, Figure 6.3d displays each German subject's SPG in the two tasks. Group C's SPG results in the RA-E task are represented by the bars. Their average SPG results in the STR E-G taskare represented by the squares.

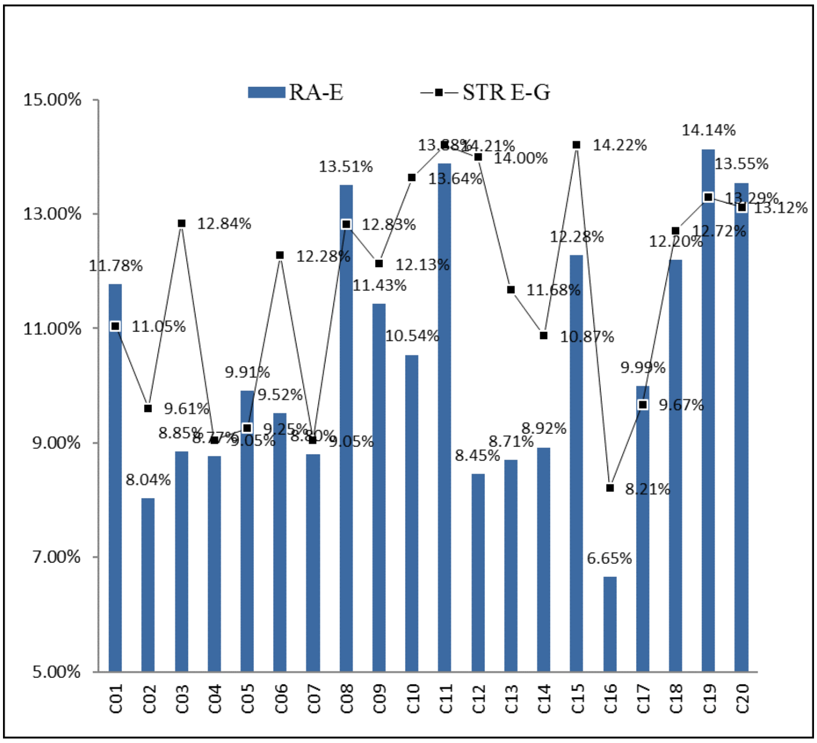


Figure 6.3d: Individual SPG (in Percentages) of Group C in RA and STR

Figure 6.3d shows that the majority (70%) of the SPG results are higher in the STR task. Then, a paired two-sample *t*-test is carried out with the SPG results obtained from the two tasks. The results are presented in Table 6.3d.

Table 6.3d: Paired Two-sample T-test for Means [SPG of Group C in RA-E and STR E-G]

T-Test: Paired Two Sample for Means

Group C	*Variable 1 (SPG in STR E-G)*	*Variable 2 (SPG in RA-E)*
Mean	0.116853768	0.104958751
Variance	0.000379988	0.000486793
Observations	20	20
Pearson Correlation	0.638788359	
Hypothesised Mean Difference	0	
df	19	
t Stat	2.986327883	
P(T<=t) one-tail	0.003793936	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	0.007587873	
t Critical two-tail	2.093024054	

Table 6.3d shows that the *t*-statistic (2.9863) is higher than the one-tail critical *t*-value (1.7291). The one-tail *p*-value (T ≤ t) is well below 0.01, suggesting that the German subjects in this experiment have a statistically higher

percentage of saccade in the STR E-G task. The standardised effect size measure of this *t*-test, the Cohen's *d*, is 0.5714. This *d* value is a *medium* effect size.

The analysis moves on to Group A’s SPG. In Figure 6.3e, the Chinese subjects’ SPG results in the RA-C task are represented by the bars. The Chinese subjects’ SPG results in the other tasks, are represented by the different black shapes: the circles for the SPG in the STR C-E task, the squares for the SPG in the STR E-C task, and the triangles for the SPG in the RA-E task.

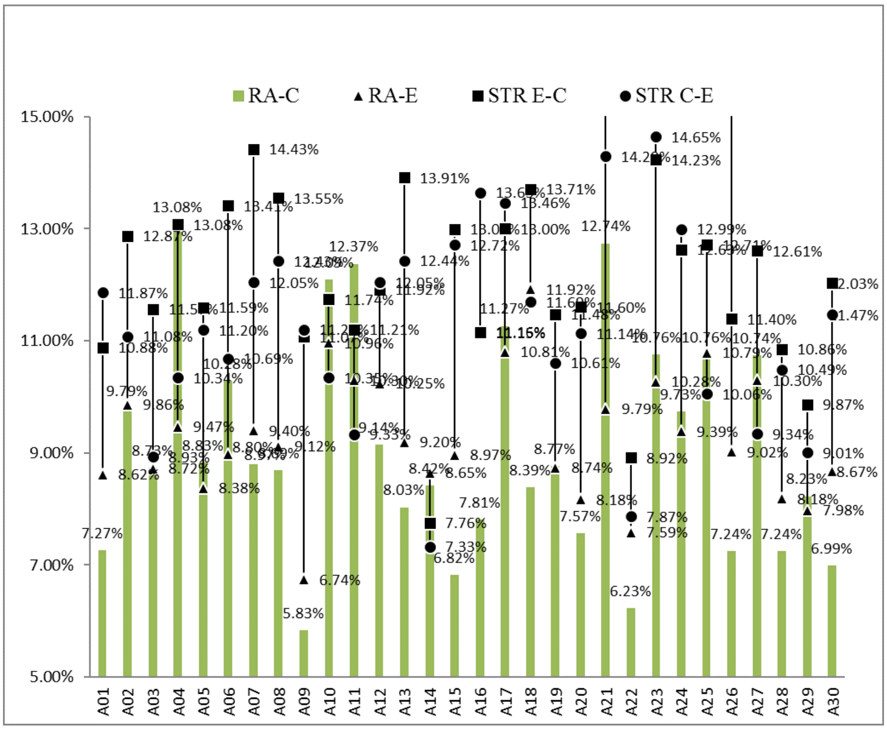


Figure 6.3e: Individual SPG (in Percentages) of Group A in RA and STR

To sum up, 80% of the Chinese subjects have a higher percentage of saccade during STR C-E than during RA-C. 93.33% of the Chinese subjects have a higher percentage of saccade during STR E-C than during RA-E.

Then, the following analysis is carried out with the SPG results from these four tasks. Assuming that the subjects in Group A have a significantly higher SPG during STR C-E than during RA-C and a significantly higher FTS during

STR E-C than during RA-E, two one-tailed and paired *t*-tests are carried out. The corresponding Null Hypotheses are $\mu 1$ (SPG in STR C-E) $\leq \mu 2$ (SPG in RA-C) (a = 0.05) and $\mu 1$ (SPG in STR C-E) $\leq \mu 2$ (SPG in RA-C) (a = 0.05). Table 6.3e and Table 6.3f present the results.

Table 6.3e: Paired Two-sample T-test for Means [SPG of Group A in STR E-C and RA-E]

T-Test: Paired Two Sample for Means

Group A	*Variable 1 (SPG in STR E-C)*	*Variable 2 (SPG in RA-E)*
Mean	0.121784442	0.093477713
Variance	0.000313138	0.000133818
Observations	30	30
Pearson Correlation	0.432903238	
Hypothesised Mean Difference	0	
df	29	
t Stat	9.440432665	
P(T<=t) one-tail	1.1986E-10	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	2.39721E-10	
t Critical two-tail	2.045229642	

Table 6.3f: Paired Two-sample T-test for Means [SPG of Group A in STR C-E and RA-C]

T-Test: Paired Two Sample for Means

Group A	*Variable 1 (SPG in STR C-E)*	*Variable 2 (SPG in RA-C)*
Mean	0.115984191	0.090885209
Variance	0.00069839	0.000380124
Observations	30	30
Pearson Correlation	-0.001350286	
Hypothesised Mean Difference	0	
df	29	
t Stat	4.183347601	
P(T<=t) one-tail	0.000121331	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	0.000242663	
t Critical two-tail	2.045229642	

It is shown that the *t*-statistics are higher than the one-tail critical *t*-values. The corresponding *p*-values are significantly lower than 0.01. The present research has an extremely high confidence level to reject the H_0. Moreover, the effect size measures of the two *t*-tests, which convey the size of an effect relative to the variability in the samples, are calculated. In the first *t*-test, the standardised effect size Cohen's *d* is 1.8932; in the second *t*-test, the Cohen's *d* is 1.0811. Both *d* values greatly exceeds the reference point for a *large* effect size. It, therefore, confirms that the SPG is statistically higher in the STR tasks than in the RA tasks.

The present study, as a result, confirms that STR tasks yield higher SPG scores than RA tasks. Since the STR tasks consume more cognitive effort than the RA tasks, it shows that a higher SPG corresponds with a larger investment of cognitive effort.

In conclusion, based on the hierarchy of the cognitive effort investment formed in Section 6.1, the relationship between SPG and cognitive effort was examined in this section. Both the descriptive and the inferential statistics presented above show that a higher SPG represents a larger amount of attention and cognitive effort devoted to the visual task. Therefore, the percentage of one's total saccade duration as a percentage of the sum of the overall saccade and fixation duration has a positive correlation with the amount of cognitive effort invested in the tasks. Since saccades indicate a shift of attention (Godijn and Theeuwes 2003: 3), having a higher SPG during STR means that the subjects' attention and cognitive resources are shifted at a higher frequency.

6.4 Summary

In this chapter, three exploratory dimensions are investigated.

The first dimension examines the cognitive investment in reading-speaking tasks based on an analysis of a series of eye-tracking measurements. Based on the results from TTT, FSD, TFD, and FC, the subjects in Group B devoted the smallest amount of cognitive effort in the RA-E task and the Chinese subjects in Group A invested the largest amount of cognitive effort in this task. Group C's cognition investment in the RA-E task was slightly higher than that of Group B, although the difference between them does not appear to be as significant as the difference between Group A and Group B. In the preliminary cross-group comparison of the three student groups' eye-tracking data, a hierarchy of their investment of cognitive effort is formed:

Group A > Group C ≥ Group B. Hence, the first hypothesis in Section 6.1 is corroborated: subjects with different L1s invest different amounts of cognitive effort in processing the same reading material during RA. In the RA-E task, the British subjects invested less cognitive effort than the non-native speakers. This study then carries out more in-depth, intra-group comparisons within Group A and Group C respectively, which aimed to examine the second hypothesis using pupil diameter, TTT, FSD, TFD, and FC. All of the indicators show a significant difference between the RA tasks and the STR tasks, no matter which experiment group was compared. Therefore, three hierarchies of the investment of cognitive effort are formed: Group A STR C-E > Group A RA-C, Group A STR E-C > Group A RA-E, Group C STR E-G > Group C RA-E. During STR, having to translate the stimulus into another language requires the subjects to engage in the more extensive sampling of the visual information and to produce the interpretation of the ST quickly. Hence, the second hypothesis is corroborated: subjects invest different amounts of cognitive effort during STR and RA. Both the Chinese subjects and the German subjects invested more cognitive effort in the STR process than in the RA process.

The second dimension examined the reliability of measuring cognitive effort using the average fixation duration. This study attempted to investigate MFD in Section 6.2 through both cross-group and intra-group comparisons. Firstly, the experimental cross-group analysis leaves us with some doubts about the use of MFD as it failed to be an accurate indicator of cognitive effort. Subsequently, this section progressed to examine this indicator by a comprehensive intra-group comparison between the MFD in the RA and the STR tasks. However, MFD was not consistent in this study. This finding makes the researcher question whether having longer fixations on average should be interpreted to mean that there has been more cognitive effort devoted to finishing the task. The eye-tracking study of Krieber *et al.* (2017) showed that eye-tracking parameters differ with respect to their consistency across different reading modes. In other words, a useful indicator applied to investigate one reading mode is not necessarily equally valuable when it is employed to examine another reading mode. Taken together, the findings of the present study suggest that MFD might not be the best measure when more than one reading modality is considered. Hence, even though this research is not suggesting researchers dismiss MFD entirely, it is asking researchers to be cautious when applying it.

The third exploratory dimension attempted to explore two new eye-tracking indicators: FTS and SPG. In Section 6.3.1, the relationship between FTS and the hierarchy of the cognitive effort investment is examined. Both descriptive and inferential statistics show that a higher FTS indicates a smaller amount of cognitive effort devoted to the visual task. This study suggests that having to transfer the text into another language reduces the FTS value because it increases the active eye movements between fixations: saccades. In Section 6.3.2, both descriptive and inferential statistics indicate that a higher SPG represents a more significant amount of attention and cognitive effort is devoted to the reading-speaking task. Since it is known that saccades indicate a shift of attention (Godijn and Theeuwes 2003: 3), having a higher SPG value during STR suggests that the subjects' attention and cognitive resources are shifted at a higher frequency in a more demanding task.

Chapter 7. Temporal EVS: A Micro Level Analysis

There has been a considerable body of spatial and temporal EVS psycholinguistic studies over the past 100 years (e.g., Buswell 1920; Laubrock and Kliegl 2015). Previous researchers have found that EVS in reading-speaking processes is inevitable (Gibson and Levin 1975). Recently, researchers interested in sight translation (STR) have studied temporal EVS during STR (e.g., Dragsted, Hansen and Sørensen 2009; Zheng and Zhou 2018). By definition, temporal EVS fits in the category of the *latency measurement* used in eye-tracking studies, which was established as "a measure of time delay, i.e., the time from the on- or offset of one event to the on- or offset of another" (Holmqvist *et al*. 2011: 428).

In this chapter, the analyses and discussions do not focus on pupil-based, fixation-based, or saccade-based measurements, which were discussed in detail in Chapter 6 (pupil diameter, TTT, FSD, TFD, and FC). Instead, temporal EVS data are applied to analyse the subjects' performance during RA and STR. The present researcher suggests that the measurement of EVS is at the *micro level* of analysis (Saldanha and O'Brien 2014: 143), as its calculation is based on single fixations that fall on specific words.

This chapter aims to investigate temporal EVS during the RA and STR process, by measuring the dynamic temporal distance between subjects' reading input and speaking output. To be more specific, the research questions are as follows:

5. Does temporal EVS correlate with major eye-tracking measurements, such as total fixation duration, fixation count, and the sum of fixation and saccade duration?

6. Is temporal EVS at sentence initials longer than temporal EVS at sentence terminals, just like spatial EVS?

7. What is the relationship between one's temporal EVS and the

cognitive effort invested in the RA and STR tasks?

8. What can we learn about temporal EVS when applying it to analyse different STR processes?

This study plans to answer these research questions by taking the following steps.

Firstly, this chapter aims to find the correlation coefficient between temporal EVS and several major eye-tracking measurements, namely total fixation duration (TFD), fixation count (FC), and the sum of fixation and saccade duration (FSD). If temporal EVS has a strong correlation coefficient with these individual eye-tracking measurements, it suggests that temporal EVS has the potential to indicate a person’s cognitive effort, although temporal EVS and these general eye-tracking indicators might represent different sub-categories of the cognitive effort devoted in a RA/STR task.

Subsequently, this research examines the temporal EVS results at the two critical positions in the meaning units. If the initial temporal EVS is longer than the terminal temporal EVS in a sentence, like spatial EVS, it indicates that temporal EVS is not a random measurement, but one that varies predictably along with cognitive processing and it can become an indicator of cognitive investment. Once the potential of temporal EVS has been discovered, the aim is to understand the relationship between the length of temporal EVS and the cognitive events of the mind. Then, the analysis moves on to the next step.

Thereafter, I aim to consider whether the length of a person’s temporal EVS in a visual-oral task correlates positively or negatively with the amount of effort he/she invested in the reading-speaking process. In Chapter 6, the amount of cognitive effort subjects devoted to these tasks was indicated by a series of eye-tracking measurements. In this chapter, the relationship between temporal EVS and cognitive effort is investigated by making an inter-group comparison among the three groups’ temporal EVS length during the RA-E task and an intra-group analysis of Group A’s and Group C’s temporal EVS during RA and STR.

Finally, temporal EVS is used to gain an insight into different STR processes: STR C-E vs. STR E-C; STR E-C vs. STR E-G. Due to the incomparability between English and Chinese languages, the present study uses RA-C and RA-E as the baseline tasks and applies temporal EVS rate of increase from RA to STR as an indicator of the increase in cognitive effort. Comparability between the Chinese and the German subjects is sought to be achieved by recruiting participants in the same age group and with similar English proficiency. To make sure that the comparison is built on comparable starting points, the RA-E tasks are used as the baseline tasks. Subjects' temporal EVS results in the RA-E task are used as the baseline, and the increase of their temporal EVS in the STR tasks represents the extra effort involved in translation. The rate of temporal EVS increase from RA to STR, therefore, is used as an indicator of the increase in cognitive effort from reading aloud to carrying out STR.

7.1 Data Presentation

Some researchers (e.g., Dragsted, Hansen, and Sørensen 2009) measured temporal EVS by *off sync fixations*: the total duration of temporal EVS is the sum of all (and only) the durations of fixations within a period of time, rather than a consecutive period from one point in time to another. The present research, however, adopts the most traditional and direct method to measure temporal EVS with the help of eye-tracking and audio-recording. This study is not suggesting that the calculation based on the duration of *off sync fixations* is any less valid, but simply pointing out that the temporal EVS calculated by the two different measures are not comparable because one is a general temporal measure whereas the other is a pure fixation-based measure. In addition, a comparison of temporal EVS results obtained by eye-tracking and off-line methods is not entirely equitable. The former measures the EVS when it is taking place naturally, whereas the latter encourages the subjects to achieve the maximum span to extract as much information as possible. If the methods used in the experiments encourage subjects to adopt different reading strategies, the results are not comparable. Chapter 5 already elaborates on how the present study measures temporal EVS, including the choice of the critical positions and other coding details. This section is a detailed presentation of the three test groups' temporal EVS at the sentence level.

Subjects in Group A completed four reading-speaking tasks: RA-E, STR E-C, RA-C, and STR C-E. Table 7.1a displays individuals' temporal EVS at sentence initials (S-I) and sentence terminals (S-T) in each task. The short descriptions following the tables present the results using summary statistics including the mean and standard deviation.

Table 7.1a: Temporal EVS (in ms) of Group A at S-I and S-T

Group A	**RA-E**		**STR E-C**		**RA-C**		**STR C-E**	
	S-I	*S-T*	*S-I*	*S-T*	*S-I*	*S-T*	*S-I*	*S-T*
A01	1052.30	742.60	1807.80	1421.00	1306.67	798.17	1791.08	1418.91
A02	1182.80	1018.30	2759.50	1082.20	1229.83	718.08	2249.83	766.42
A03	1313.10	1102.80	1204.70	1201.50	1278.92	720.25	1605.58	1158.00
A04	1106.30	907.40	1628.00	1013.00	992.92	765.08	1592.33	956.09
A05	1122.10	886.56	1844.60	1406.50	1283.25	786.67	3319.67	1054.82
A06	1151.10	824.60	2268.20	1382.40	1197.33	667.25	3274.78	921.00
A07	1103.20	787.80	2114.50	1185.20	1196.00	759.33	2037.08	987.33
A08	1169.50	878.80	1740.00	1505.20	1181.00	813.25	1768.17	1061.25
A09	1338.50	904.67	2871.00	2027.00	1115.00	625.25	2041.75	1507.00
A10	1104.80	842.30	2614.20	1351.80	1550.25	797.67	2465.67	1283.30
A11	1134.10	980.60	1913.00	1487.67	1098.42	762.50	1601.42	1509.00
A12	1166.20	866.70	2033.30	1742.44	1233.33	717.25	1396.42	927.10
A13	1130.80	796.50	1662.00	1593.22	1207.58	658.50	1595.75	925.00
A14	1202.00	907.90	1774.20	1375.20	1342.58	844.75	1524.00	1109.73
A15	1156.10	886.30	2251.00	1734.80	1044.75	575.83	1813.50	1100.25
A16	1632.90	811.40	1924.90	1841.00	1594.83	573.50	2348.92	1323.36
A17	1247.40	805.90	1783.50	1354.10	1310.42	837.58	1506.83	1257.92
A18	1112.00	819.90	1706.80	1448.56	1094.17	750.92	1503.60	1361.00
A19	1069.20	832.70	3866.00	2381.63	1200.67	717.42	2529.00	1850.42
A20	1189.50	852.78	2085.50	1689.75	1105.67	636.58	1597.58	1132.67
A21	1331.90	938.90	1740.60	1212.30	1241.17	688.33	1569.67	1619.75
A22	1152.90	823.20	3397.70	1558.00	1090.58	701.58	1768.50	1477.45
A23	1137.50	783.40	2078.40	1812.13	1158.08	715.75	1370.67	1204.50
A24	1176.00	918.60	1955.30	1763.00	1174.00	1095.08	1786.33	1440.17
A25	1170.20	871.90	1974.70	1877.00	1077.08	670.67	1751.08	1133.00
A26	1365.80	889.60	3700.10	2018.50	1171.25	759.92	2862.09	1525.64
A27	1037.10	797.30	1864.56	1232.10	1023.00	709.33	1782.91	1495.30
A28	1206.10	802.30	1647.90	1319.00	991.00	699.58	1540.58	824.92
A29	1139.30	809.00	2089.40	1821.56	1100.25	622.75	1463.17	740.18
A30	1048.50	757.90	2693.50	1941.75	1024.08	552.25	1815.58	1339.09

Table 7.1a displays the duration of the Chinese subjects' temporal EVS at the sentence level in the four tasks they completed. To sum up, the average length of Group A's temporal EVS is 1181.64 ms (n = 30, *SD* = 119.35) at the sentence initials and 861.62 ms (n = 30, *SD* =78.19) at the sentence terminals in the task RA-E; 2166.50 ms (n = 30, *SD* = 623.72) at the sentence initials and 1559.32 ms (n = 30, *SD* = 320.64) at the sentence terminals in the task STR E-C; 1187.14 ms (n = 30, *SD* = 142.11) at the sentence initials and 724.70 ms (*SD* = 103.28) at the sentence terminals in the task RA-C; 1909.12 ms (n = 30, *SD* = 521.52) at the sentence initials and 1213.69 ms (n = 30, *SD* = 273.67) at the sentence terminals in the task STR C-E. The *SD* values of temporal EVS in the two STR tasks, for instance the *SD* of temporal EVS at the sentence initials in the task STR E-C, are generally higher. This result is probably due to a larger interpersonal difference among subjects during STR. The *SD* values of temporal EVS in the two RA tasks are generally lower because of a smaller interpersonal difference among subjects' reading performance.

Figure 7.1a displays individuals' temporal EVS at S-I and S-T in RA-E and STR E-C.

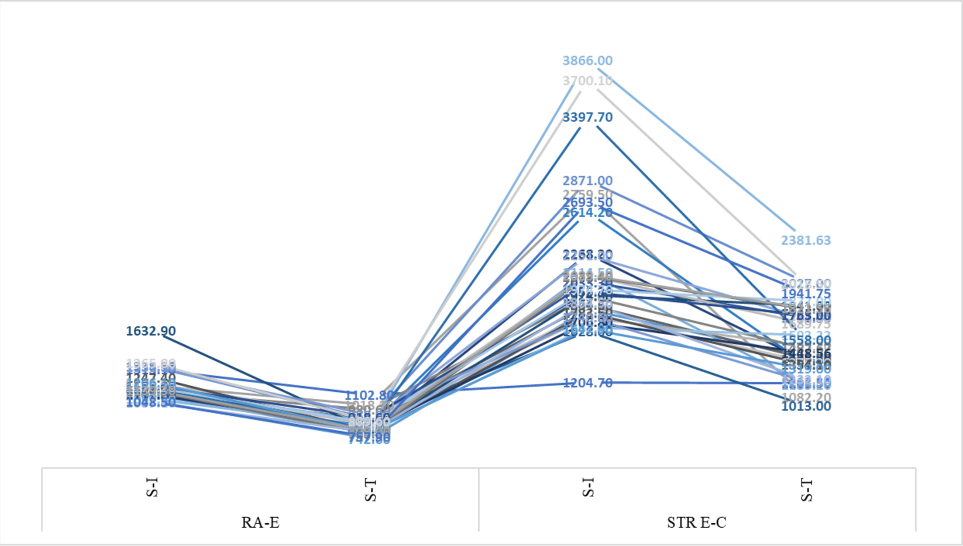


Figure 7.1a: Temporal EVS (in ms) of Group A at S-I and S-T in RA-E and STR E-C

Figure 7.1b displays individuals' temporal EVS at S-I and S-T in RA-C and STR C-E. RA-C and STR C-E are shown in the same figure because these two tasks have the same ST.

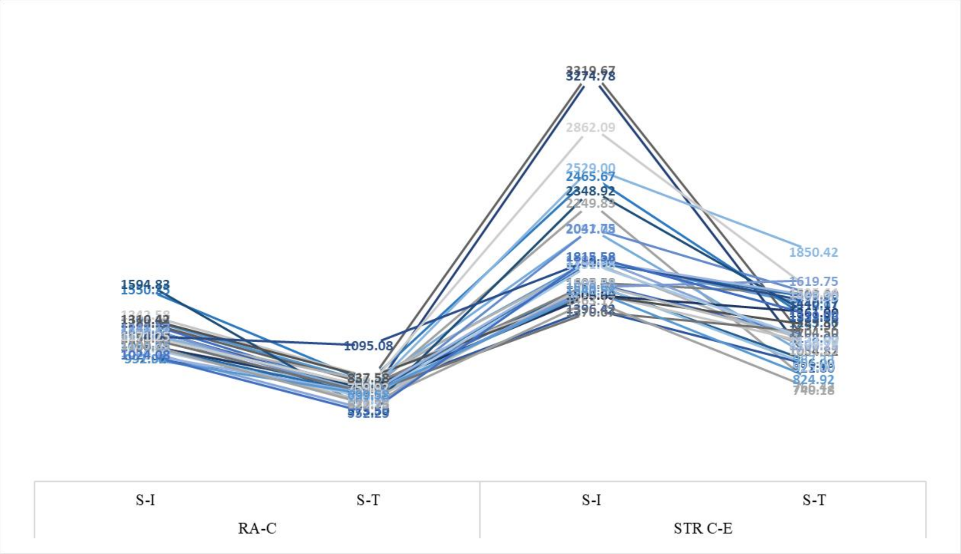


Figure 7.1b: Temporal EVS (in ms) of Group A at S-I and S-T in RA-C and STR C-E

Figure 7.1a and Figure 7.1b show a relatively interesting contrast between the different data sets. For instance, subjects have a wider range of temporal EVS in STR tasks whereas temporal EVS in RA tasks has a smaller difference among subjects. The maximum temporal EVS in Group A occurs at S-I in STR E-C (3866.00 ms) and the minimum temporal EVS in Group A occurs at S-T in RA-C (552.25 ms).

Subjects in Group B completed one reading-speaking task: RA-E. Table 7.1b displays individuals' temporal EVS at sentence initials (S-I) and sentence terminals (S-T) in this task. The short descriptions following the tables present the results using summary statistics including the mean and the standard deviation.

Table 7.1b: Temporal EVS (in ms) of Group B at S-I and S-T

Group B	**RA-E**	
	S-I	*S-T*
B01	972.90	745.30
B02	1034.00	1011.90
B03	1052.50	722.44
B04	982.40	729.40
B05	1031.20	700.20
B06	1214.30	1217.30
B07	892.70	762.30
B08	1040.20	589.60
B09	875.50	823.70
B10	822.90	825.70
B11	1117.90	620.10
B12	1299.30	1038.40
B13	1173.60	788.00
B14	890.20	665.50
B15	1041.70	773.60
B16	1037.20	844.40
B17	1010.40	959.30
B18	1114.30	839.30
B19	1196.10	968.10
B20	1008.50	636.70
B21	984.30	766.30
B22	920.60	1052.40
B23	991.90	651.10
B24	1094.60	923.30
B25	932.30	710.50
B26	1294.70	1125.40
B27	1065.50	971.40
B28	927.60	473.50
B29	942.90	851.30
B30	1086.10	928.80
B31	1105.10	870.00
B32	1068.80	586.01
B33	880.90	1041.80
B34	1108.90	1244.10
B35	1067.50	844.50

Table 7.1b displays the British subjects' temporal EVS lengths at the sentence level during the RA-E task. The average length of Group B's temporal EVS is 1036.56 ms (n = 35, *SD* = 113.65) at the sentence initials; 837.19 ms (n = 35, *SD* = 180.94) at the sentence terminals in the task RA-E. Figure 7.1c display individuals' temporal EVS at S-I and S-T in RA-E. Their *SD* values at the terminal position is higher than at the initial position, indicating that there is a smaller interpersonal difference when reading started at sentence initials. As the subjects reached terminal positions in the text, some reduced

their temporal EVS much more than others.

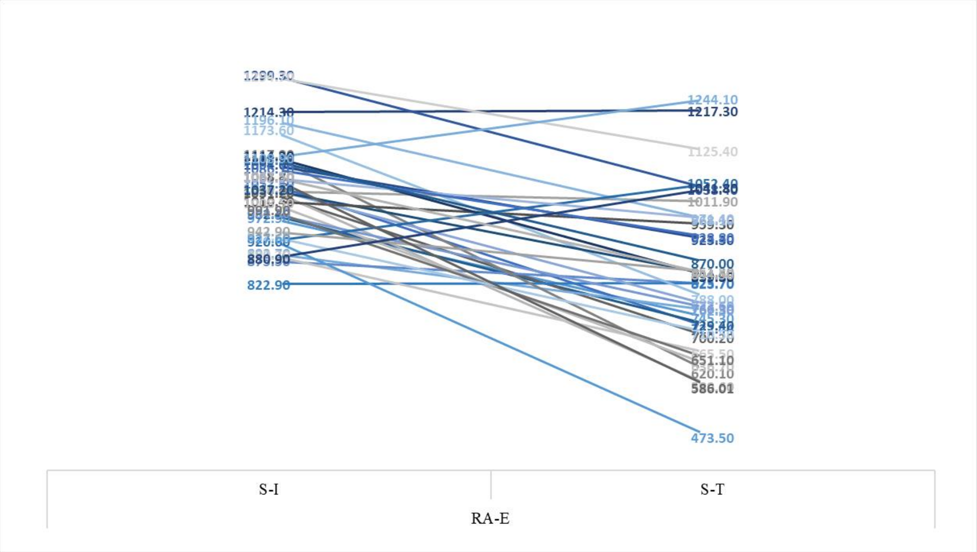


Figure 7.1c: Temporal EVS (in ms) of Group B at S-I and S-T in RA-E

Figure 7.1c does not show a very clear contrast between temporal EVS at S-I and S-T. Nevertheless, generally, individuals in Group B seem to have longer average temporal EVS at S-I in RA-E. The maximum temporal EVS in Group B occurs at S-I (1299.30 ms) and the minimum temporal EVS occurs at S-T (473.50 ms).

Subjects in Group C completed two reading-speaking tasks: RA-E and STR E-G. Table 7.1c displays individuals' temporal EVS at sentence initials (S-I) and sentence terminals (S-T) in both tasks. The short descriptions following the tables present the results using summary statistics including the mean and the standard deviation.

Table 7.1c: Temporal EVS (in ms) of Group C at S-I and S-T

Group C	**RA-E**		**STR E-G**	
	S-I	*S-T*	*S-I*	*S-T*
C01	1023.67	746.50	2505.80	1850.00
C02	934.00	682.20	2112.30	1653.00
C03	1247.80	1115.40	3851.10	2692.90
C04	1105.20	752.60	2531.80	1513.50
C05	1066.60	772.10	3536.70	2444.40
C06	1036.90	856.80	2686.00	1948.75
C07	935.70	905.10	2388.90	1986.89
C08	1063.50	822.30	1292.10	1045.70
C09	1067.80	971.90	1847.90	1515.00
C10	1112.10	906.10	3004.10	2049.60
C11	946.10	671.60	2350.50	1954.78
C12	1109.80	862.30	1878.30	1871.22
C13	1156.50	770.20	2271.30	2204.40
C14	1245.90	974.30	2471.80	2281.67
C15	1100.70	786.70	2465.50	2358.11
C16	930.10	687.30	2735.00	2200.44
C17	1348.10	1131.00	2470.00	1688.40
C18	911.40	723.30	1737.00	1971.90
C19	1206.10	862.80	3614.20	2349.63
C20	1159.30	926.60	1672.50	1653.50

Table 7.1c displays the German subjects' individual temporal EVS at the sentence level in the two tasks. The average length of Group C's temporal EVS is 1085.36 ms (n = 20, *SD* = 120.22) at the sentence initials and 846.35 ms (n = 20, *SD* = 132.14) at the sentence terminals in the task RA-E; 2471.14 ms (n = 20, *SD* = 658.86) at the sentence initials and 1961.69 ms (n = 20, *SD* = 385.76) at the sentence terminals in the task STR E-G. Figure 7.1d display individuals' temporal EVS at S-I and S-T in both tasks. Like Chinese subjects' *SD* values of temporal EVS, German subjects' *SD* values of temporal EVS in the STR task is also higher. Again, the present study suspected that it is due to a larger interpersonal difference among subjects' STR competence and characteristics.

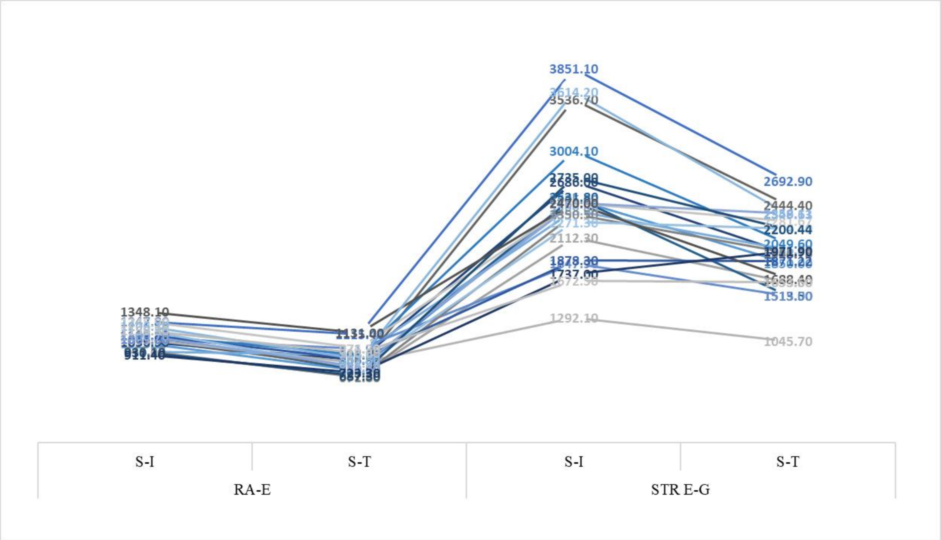


Figure 7.1d: Temporal EVS (in ms) of Group C at S-I and S-T in RA-E and STR E-C

Like Figure 7.1a and Figure 7.1b, Figure 7.1d also shows a relatively clear contrast between different data sets. The subjects have a wider range of temporal EVS in STR E-G whereas temporal EVS in RA-E have a smaller difference among subjects. The maximum temporal EVS in Group C occurs at S-I in STR E-G (3851.10 ms) and the minimum temporal EVS occurs at S-T in RA-E (671.60 ms).

The series of figures present some interesting patterns, such as some contrast between temporal EVS at S-I and S-T and some differences between temporal EVS in different tasks. These preliminary patterns are tested and discussed in detail in the coming sections. Before that, however, the present study briefly touches upon the relationship between temporal EVS and other variables in RA, in the hope of improving our understanding of temporal EVS. This preliminary examination is carried out with temporal EVS and several additional variables in RA-E because the use of EVS as an indicator originated in English reading studies.

As stated in detail in Chapter 5, the reading texts used in RA tasks and the source texts used in STR tasks consist of sentences of varying length. Therefore, whether the word count of these sentences has a certain effect on the duration of a subjects' temporal EVS is analysed here. In brief, the correlation coefficients between subjects' temporal EVS at S-I and S-T of all the sentences and the individual word count of all the sentences are -0.0469 and -0.2274. This result shows that although the sentences in the reading text/ST are different lengths, their word count does not have a strong influence on the duration of one's temporal EVS in RA-E in this experiment, no matter if the subjects are Chinese, British, or German. It is therefore suggested that using authentic reading texts that consist of sentences of varying word count is not an issue when studying temporal EVS.

In general, researchers believe that the faster the reading speed, the wider the spatial EVS (e.g. Buswell 1920; De Luca *et al*. 2013). Hence, there is this argument that the wider the spatial EVS, the better the reader. The present study, hopes to test this theory with the empirical evidence collected when using state-to-art research methods. Moreover, whether the duration of temporal EVS also increases with reading speed has not been explored in previous studies. Hence, the present study also aims to carry out a preliminary investigation on the relationship between temporal EVS at S-I and S-T in RA-E and these two variables: reading speed and spatial EVS.

Total task time (TTT) is how much time (in seconds) the subjects take to complete the same RA-E task. The reading material consists of 200 English words, so the subjects' reading speed can be calculated by: 200 words/total task time. The average reading speed of Group A is 2.18 words/second (n = 30, *SD* = 0.22); the average reading speed of Group B is 3.05 words/second (n = 35, *SD* = 0.38); and the average reading speed of Group C is 2.83 words/second (n = 35, *SD* = 0.24). The groups' average reading speed suggests that the British subjects read at a higher speed than other subjects during RA-E.

Meanwhile, the average width of Group A's spatial EVS is 3.05 words (n = 30, *SD* = 0.22) at S-I and 2.20 words (n = 30, *SD* = 0.16) at S-T; the average width of Group B's spatial EVS is 3.75 words (n = 30, *SD* = 0.32) at S-I and 2.54 words (n = 30, *SD* = 0.35) at S-T; the average width of Group C's spatial

EVS is 3.61 words (n = 30, *SD* = 0.26) at S-I and 2.55 words (n = 30, *SD* = 0.21) at S-T in the task RA-E. The group who read at the highest speed has the widest spatial EVS at S-I in RA-E. Almost a hundred years ago, Buswell (1921) applied spatial EVS when investigating reading competence and suggested that developing reading competence correlates with expanding one's spatial EVS. Also, according to De Luca *et al*. (2013: 11), a wider spatial EVS is closely associated with faster reading, whereas a narrower spatial EVS is associated with slower reading. The present study's findings regarding spatial EVS is generally in line with the findings of other researchers who investigated the differences in spatial EVS between good and poor readers in the last century (e.g., Buswell 1921). Hence, the present study acknowledges the findings of the previous researchers who argued that the width of spatial EVS has a "consistent and robust" relationship with one's reading ability and can be used to judge one's reading skill (Levin and Addis 1979: 53). These findings encouraged me consider if temporal EVS could be the same kind of indicator. In other words, is there also such a relationship between reading speed and the length of temporal EVS?

As presented in a previous paragraph, the average duration of Group A's temporal EVS is 1181.64 ms (n = 30, *SD* = 119.35) at S-I and 861.62 ms (n = 30, *SD* =78.19) at S-T; the average duration of Group B's temporal EVS is 1036.56 ms (n = 35, *SD* = 113.65) at S-I and 837.19 ms (n = 35, *SD* = 180.94) at S-T; the average duration of Group C's temporal EVS is 1085.36 ms (n = 20, *SD* = 120.22) at S-I and 846.35 ms (n = 20, *SD* = 132.14) at S-T in the task RA-E. On average, Group A has the longest temporal EVS at both S-I and S-T, whereas Group B has the shortest temporal EVS at both S-I and S-T in RA-E. Since Group B has a shorter temporal EVS than the other two groups at both the sentence initials and the sentence terminals during RA-E, a higher reading speed is thus associated with a shorter temporal EVS. This finding regarding temporal EVS and reading speed is in line with some findings in recent years. For example, Dragsted and Hansen (2008) found that difficult words cause longer temporal eye-key spans than easy words. Timarová *et al.* (2015) detected a shorter temporal EVS for more experienced interpreters. In another study viewing the eye-key span as a measure of translation complexity, Carl and Schaeffer (2016) supported Dragsted's (2010) findings that professional translators have a shorter temporal eye-key span than student translators. As shown in chapter 6, subjects in Group B are faster readers compared to the other two groups. Furthermore, this chapter has shown that subjects in Group B have a shorter temporal EVS compared to the other two groups. Based on previous research and these preliminary results, the present study concludes that there could be a correlation between reading faster and having a shorter temporal EVS. Perhaps faster readers with the current level of reading competence synchronised the text processing, speech production, and speech monitoring to a greater degree, whereas slower readers in this study are less capable of doing so.

When I considered the length of temporal EVS and the width of spatial EVS, another interesting pattern surfaced. The group that has the shortest temporal EVS has the widest spatial EVS, and the group which has the longest temporal EVS has the narrowest spatial EVS at both S-I and S-T in RA-E. This negative correlation, however, is not surprising because as stated in Chapter 3, temporal EVS and spatial EVS are two related but completely different measurements. While a wider spatial EVS can be seen as a sign of greater reading competence, a longer temporal EVS does not necessarily point to the same conclusion. Earlier in this century, Laubrock and Kliegl (2015) discovered that the duration of temporal EVS at the end of a fixation is positively correlated with the likelihood of having regressions and refixations and argued that the regressions and refixations are means to regulate temporal EVS when it becomes too large. In essence, Laubrock and Kliegl (2015: 2) suggested that "readers do not want the eyes to go too far ahead of the voice". This shows that having a substantially longer temporal EVS might indicate cognitive demand. The opposite patterns in previous research and the differences found by the present study both show that the length of temporal EVS and the width of spatial EVS are not positively correlated. The importance of this finding is that researchers should be aware of the difference between temporal EVS and spatial EVS in future studies. As proposed earlier in this thesis, temporal EVS and spatial EVS should not be confused.

Overall, temporal EVS correlates negatively with reading speed and it does not expand positively with spatial EVS. Subsequently, an investigation of temporal EVS was conducted because the present researcher believes that the temporal dynamics between one's eyes and voice reflect a constellation of the sub-processes that underpin reading. Essentially, as a buffering mechanism, temporal EVS represents a time lag between the reading comprehension and oral production. During visual tasks, such as RA and STR, temporal EVS has to be expanded because it serves to remedy the situation, when "higher level processes lag behind as new visual information is accumulated" (Slowiaczek 1983: 348). In other words, if the higher-level processes could be carried out relatively on time, less new visual information is accumulated and left unprocessed. This could potentially be the reason why faster readers have a shorter temporal EVS than slower speakers. The next section, begins with an investigation of the potential of temporal EVS as an indicator of cognitive processing during reading-speaking coordination.

7.2 The Potential of Temporal EVS as an Indicator of Cognitive Effort

After the data coding and presentation, this section analyses temporal EVS at the sentence level, in the hope of discovering its general characteristics.

7.2.1 Temporal EVS and Major Eye-tracking Indicators

In Chapter 6, the relationship between cognitive effort and eye-tracking indicators, such as total fixation duration and fixation count, were discussed. The aim of this section (7.2.1) is to shed some light on the relationship between temporal EVS and these major fixation-based eye tracking indicators that have been used not only in the present study but also by other researchers (e.g., Holmqvist *et al*. 2011; Liu, Zheng, and Zhou 2018).

The underlying differences between temporal EVS and the major fixation-/saccade-based indicators used in the present study are threefold. First and foremost, the calculation and analysis of general eye-movement data is at the *macro level* as it is based on large segments of the eye-tracking recordings and an analysis of eye-tracking data within relatively longer AOIs, whereas the presentation and analysis of temporal EVS during reading and STR is at a *micro level* based on individual fixations and syllables. Secondly, the eye-tracking indicators used in the present study are relatively long-term measures that indicate all the events during a specific period. Temporal EVS, on the other hand, is an instantaneous measure of the gap in time between a person's reading input and speaking output. Thirdly, while presenting the other indicators based solely on the eye-tracking data, calculating the length of the temporal EVS cannot be completed without audio data. While the other fixation- or saccade-based indicators can be calculated directly, the execution of temporal EVS requires manual labour as the eye-tracking software is currently not capable of this task.

Because of all these differences, a hasty generalisation should not be made about macro level and micro level measures. Nevertheless, the present study proposes a preliminary examination of the correlation coefficient between the duration of all subjects' temporal EVS and major eye-tracking measurements in all the reading-speaking tasks carried in the experiment in the hope of establishing the potential of temporal EVS as a new indicator of cognitive effort. Table 7.2a shows the Pearson's correlation coefficient (*r*) between temporal EVS at both S-I and S-T and three major eye-tracking measurements.

Table 7.2a: Pearson's Correlation Coefficient (*r*) between Temporal EVS at S-I and S-T and TFD, FC, and FSD

r	TFD	FC	FSD
Temporal EVS at S-I	0.782496	0.774857	0.785984
Temporal EVS at S-T	0.731047	0.746454	0.741600

The Pearson's correlation coefficient between temporal EVS and the other three measurements is considered strong for the purposes of this study. By a small amount, temporal EVS correlates most strongly with FSD, which is a fixation-saccade based eye-tracking indicator. The strong correlations here suggest that temporal EVS has the potential to indicate one's cognitive effort, although temporal EVS and these eye-tracking indicators are perhaps representatives of different sub-categories of cognition during reading-speaking. Hence, the present study hypothesises that temporal EVS, like TFD, FC, and FSD, is also capable of indicating the level of cognitive effort devoted during RA and STR.

In summary, temporal EVS is found mostly to be in line with three of the major eye-tracking indicators regarding the investment of cognitive effort. In fact, temporal EVS has a strong correlation coefficient with each of them at both S-I and S-T in both RA and STR tasks. Temporal EVS at sentence initials and sentence terminals are compared in the next section. If the initial temporal EVS is longer than the terminal temporal EVS in a sentence, it suggests that the length of temporal EVS during RA and STR has the potential to be an indicator of cognitive investment in eye-tracking based RA/STR studies.

7.2.2 Temporal EVS at S-I and S-T

In this section, the study presents an intra-group comparison within Group A, Group B, and Group C, aiming at examining their temporal EVS at sentence initials and sentence terminals in both the RA tasks and the STR tasks. The intra-group comparison of the temporal EVS at sentence level takes place within one test group and one task each time. To be specific, Group A's temporal EVS results at the sentence initials and the sentence terminals in the four tasks (RA-C, RA-E, STR C-E, and STR E-C) are contrasted with each other. Group B's temporal EVS results at the sentence initials and the sentence terminals in the one RA task (RA-E) are compared. Group C's temporal EVS results at the sentence initials and the sentence terminals in the two tasks (RA-E, STR E-G) are contrasted respectively. To illustrate, the length of Group A's temporal EVS at the sentence initials in STR E-C is contrasted only with the length of their own temporal EVS at the sentence terminals in the same task. No cross-group or cross-task comparison is made in this section.

The first intra-group comparison between temporal EVS at the sentence initials and the sentence terminals is made within Group B, whose subjects only completed one reading aloud task in English.

Both the descriptive statistics and the inferential statistics are presented in Table 7.2b.

Table 7.2b: Temporal EVS (in ms) of Group B at S-I and S-T in RA-E

Group B	RA-E	
Position	S-I	S-T
μ	1036.56	837.19
SD	113.56	180.94
p (T≤t) one tail	5.82338E-17 < 0.01	
Cohen's d	1.32 > 0.8	
*H_0	μ1 (EVS at S-I)≤μ2 (EVS at S-T) (a = 0.05)	

The average length of Group B's temporal EVS is 1036.56 ms (n = 35, *SD* = 113.56) at the sentence initials and 837.19 ms (n = 35, *SD* = 180.94) at the sentence terminals in the RA-E task. At first glance, the British subjects in Group B do have a longer temporal EVS at the sentence initials on average, but this research still sought evidence from inferential statistics. A *t*-test (paired two-sample for means) is conducted. Details of this *t*-test are presented by a table in the appendices (see Appendix 4, Table 7.2b*). Assuming that the Group B subjects have a significantly longer EVS at the sentence initials during the RA-E task, the Null Hypothesis is μ1 (EVS at S-

I) $\leq \mu 2$ (EVS at S-T) (a = 0.05). The *t*-statistic (6.9557) is much higher than the one-tail critical *t*-value (1.6909). With 35 subjects in Group B, the significant difference is found to be at the $p < 0.01$ level. The present research has an extremely high confidence level to reject the H_0 as the extremely low *p*-value ($p < 0.01$) in the one-tailed *t*-test indicates that the group's average result is reliable. Besides, the Cohen's *d* measure for this *t*-test, which is the standardised effect size that conveys the size of an effect relative to the variability in the samples, is calculated. Cohen's $d = (M_2 - M_1) / SD_{\text{pooled}}$; $SD_{\text{pooled}} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. The result is a very *large* effect size (1.3199) (reference points for small, medium, and large effects are 0.2, 0.5, and 0.8 respectively). In brief, the results show that the Group B's temporal EVS is significantly longer at S-T than at S-I.

The second comparison is made within Group C. It is made between the length of the temporal EVS at the sentence initials and the sentence terminals during RA-E and STR E-G. Both the descriptive statistics and the inferential statistics are presented in Table 7.2c.

Table 7.2c: Temporal EVS (in ms) of Group C at S-I and S-T in RA-E

Group C	RA-E		STE E-G	
Position	S-I	S-T	S-I	S-T
μ	1085.36	846.35	2471.14	1961.69
SD	120.22	132.14	658.86	385.76
p (T≤t) one tail	7.42355E-11 < 0.01		2.24949E-05 < 0.01	
Cohen's *d*	1.89 > 0.8		0.94 > 0.8	
*H_0	µ1 (EVS at S-I)≤µ2 (EVS at S-T) (a = 0.05)			

The average length of Group C's temporal EVS at the sentence initials (n = 20, μ = 1085.36 ms, *SD* = 120.22) is longer than that at the sentence terminals (n = 20, μ = 846.35 ms, *SD* = 132.14) in the task RA-E. It is also longer at the sentence initials (n = 20, μ = 2471.14 ms, *SD* = 658.86) and shorter at the sentence terminals (n = 20, μ = 1961.69 ms, *SD* = 385.76) in the task STR E-G. So, on average, Group C's temporal EVS at the sentence initials are longer than that at the sentence terminals. Again, *t*-tests are carried out to check if the difference between the temporal EVS at the sentence initials and the sentence terminals is statistically significant. Details of these *t*-tests are presented by a table in the appendices (see Appendix 4, Table 7.2c*). Assuming that the subjects in Group C have a significantly longer temporal EVS at the sentence initials during both the RA-E and the STR tasks, the Null Hypothesis is $\mu 1$ (EVS at S-I) $\leq$ $\mu 2$ (EVS at S-T) ($a = 0.05$). Two *t*-tests (paired two-sample for means) are conducted. The *t*-statistics (12.4007; 5.2563) are much higher than the one-tail critical *t*-values (1.7291; 1.7291). With 20 subjects in Group C, the significant difference is found to be at the $p < 0.01$ level. The H_0 is thus rejected with a high confidence level. The present researcher believes that Group C's temporal EVS at the sentence initials is statistically longer than that at the sentence terminals in both the RA and the STR tasks. Furthermore, in the first *t*-test, Cohen's $d = 1.8921$; in the second *t*-test, Cohen's $d = 0.9437$. Both results suggest that the effect sizes are quite *large*, because 0.8 is considered the reference point for a *large* effect size. In brief, the results show that the Group C's temporal EVS is significantly longer at S-T than at S-I.

Group A's data is examined subsequently in this section. Group A is the group that has a relatively large number of subjects who took part in four tasks. Both the descriptive statistics and the inferential statistics are presented in Table 7.2d.

Table 7.2d: Temporal EVS (in ms) of Group A at S-I and S-T in RA and STR

Group A	RA-E		STE E-C	
Position	S-I	S-T	S-I	S-T
μ	1181.64	861.62	2166.50	1559.32
SD	119.35	78.19	623.72	320.64
p (T≤t) one tail	4.97227E-15 < 0.01		1.94508E-07 < 0.01	
Cohen's *d*	3.16 > 0.8		1.22 > 0.8	
*H_0	μ1 (EVS at S-I)≤μ2 (EVS at S-T) (a = 0.05)			
Group A	RA-C		STE C-E	
Position	S-I	S-T	S-I	S-T
μ	1187.14	724.70	1909.12	1213.69
SD	142.11	103.28	521.52	273.67
p (T≤t) one tail	4.15827E-16 < 0.01		1.11843E-07 < 0.01	
Cohen's *d*	3.72 > 0.8		1.67 > 0.8	
*H_0	μ1 (EVS at S-I)≤μ2 (EVS at S-T) (a = 0.05)			

Descriptively, the average length of Group A's temporal EVS is longer at the sentence initials (n = 30, μ = 1181.64 ms, *SD* = 119.35) than at the sentence terminals (n = 30, μ = 861.62 ms, *SD* = 78.19) in the task RA-E; longer at the sentence initials (n = 30, μ = 2166.50 ms, *SD* = 623.72) than it is at the sentence terminals (n = 30, μ = 1559.32 ms, *SD* = 320.64) in the task STR E-C; longer at the sentence initials (n = 30, μ = 1187.14 ms, *SD* = 142.11) than that at the sentence terminals (n = 30, μ= 724.70 ms, *SD* = 103.28) in the task RA-C; longer at the sentence initials (μ = 1909.12 ms, *SD* = 521.52) and shorter at the sentence terminals (n = 30, μ = 1213.69 ms, *SD* = 273.67) in the task STR C-E.

In short, Group A's temporal EVS at the sentence initials is longer than that at the sentence terminals regardless of the task. Then, *t*-tests are carried out to check if the difference between the temporal EVS at the sentence initials and terminals is statistically significant. Details of these *t*-tests are presented in a table in the appendices (see Appendix 4, Table 7.2d*). Assuming that the Group A subjects have a significantly longer EVS at the sentence initials during both the RA and the STR tasks, the Null Hypothesis is $\mu 1$ (EVS at S-I) ≤ $\mu 2$ (EVS at S-T) (a = 0.05). All the *t*-statistics are a lot higher than the one-tail critical *t*-values. With the 30 subjects' results, the significant difference is found to be at the $p < 0.01$ level. The extremely low *p*-values ($p < 0.01$) in all four *t*-tests indicate that the inferential data is in line with the descriptive statistics. Furthermore, the effect size measures of the *t*-tests are calculated: RA-E 3.1637, STR E-C 1.2244, RA-C 3.7227, STR C-E 1.6699. The four Cohen's *d* measures suggest that the effect sizes of these paired *t*-tests well exceed the reference point for a *large* effect size. The present research, therefore, has an extremely high confidence level to conclude that Group A's temporal EVS at the sentence initials is always longer than that at the sentence terminals during both the RA and the STR tasks.

This section is dedicated to investigating whether the relationship between the length of temporal EVS during reading and the position in a sentence corresponds to what the researchers Quantz (1897) and Buswell (1920) stated about spatial EVS: it is longer at the sentence initials than at the sentence terminals. The findings suggest that the subjects from different language backgrounds have a statistically longer temporal EVS at the sentence initials than at the sentence terminals in both RA tasks and STR tasks. The present study, therefore, argues that since the duration of a subject's temporal EVS varies within a sentence, the temporal EVS is not a random measurement, but one that varies predictably and potentially varies due to influences from cognitive processing. Readers/translators shorten their temporal EVS at the sentence terminals because of their anticipation of the end of a meaning unit. Moreover, the sentence unit turns out to be a suitable unit for studying temporal EVS in both RA and STR because a statistically significant difference between the length of temporal EVS at these two positions in a sentence is found in both tasks.

In conclusion, in this section comparisons are made to see how the systematic changes in the independent variable "the location in sentences" affects the dependent variable "the length of temporal EVS". It is shown that the readers/translators tend to slow down their oral production at sentence initials to accumulate as much information as possible. As a result, one's temporal EVS is longer at sentence initials. Towards the end of the sentence, subjects prepare to wrap up the cognitive processing of this sentence by speeding up their oral production. As their voice catches up with their eyes, their temporal EVS is reduced. The present research, as a result, suggests that the length of temporal EVS taking place during RA and STR has the potential to be used as a measurement of cognitive effort in our eye-tracking based empirical study. As an indicator, it might be able to show researchers the amount of cognitive resources the readers devoted to completing different reading-speaking tasks.

7.3 The Correlation between Temporal EVS Length and Cognitive Effort

The previous section (Section 7.2) suggests that the length of temporal EVS during RA and STR has the potential to become an indicator of the amount of cognitive resources readers devote during different reading tasks. It is now important to map out the correlation between temporal EVS and cognitive effort, which is presented in Chapter 6, independently of the discussion related to temporal EVS. Hence, the aim of this section (Section 7.3) is to establish the relationship between temporal EVS length and the amount of cognitive effort invested during RA and STR by taking the following two steps:

1) making an inter-group comparison among three groups' RA-E,
2) making an intra-group comparison between RA and STR.

7.3.1 Temporal EVS of Group A, B, and C in RA-E

In this section, the study carries out a cross-group examination of temporal EVS of Group A, B, and C in the RA-E task, with a view to gaining an insight into the characteristics of these three groups' temporal EVS at the sentence level.

The Chinese subjects in Group A, the British subjects in Group B, and the German subjects in Group C completed the same reading task: reading aloud in English (RA-E). In Chapter 6, a hierarchy of their investment of cognitive effort (Group A > Group C ≥ Group B) was formed using the conventional indicators including Total Task Time (TTT), Total Fixation Duration (TFD), Fixation Count (FC), and Pure Gaze Activities (PGA), which is also known as the sum of Fixation and Saccade Duration (FSD). This hierarchy is viewed as a basis for the analysis in this section.

The average temporal EVS sizes are compared to consider the native English speakers' and non-native English speakers' eye-voice coordination at the sentence level during RA-E to map out the relationship between temporal EVS and cognitive effort invested in the task.

First, two sets of box-and-whisker plots are drawn in preparation for the subsequent analyses. In Figure 7.3a, the length (in ms) of Group A's, B's, and C's temporal EVS at the sentence initials and terminals in RA-E is displayed.

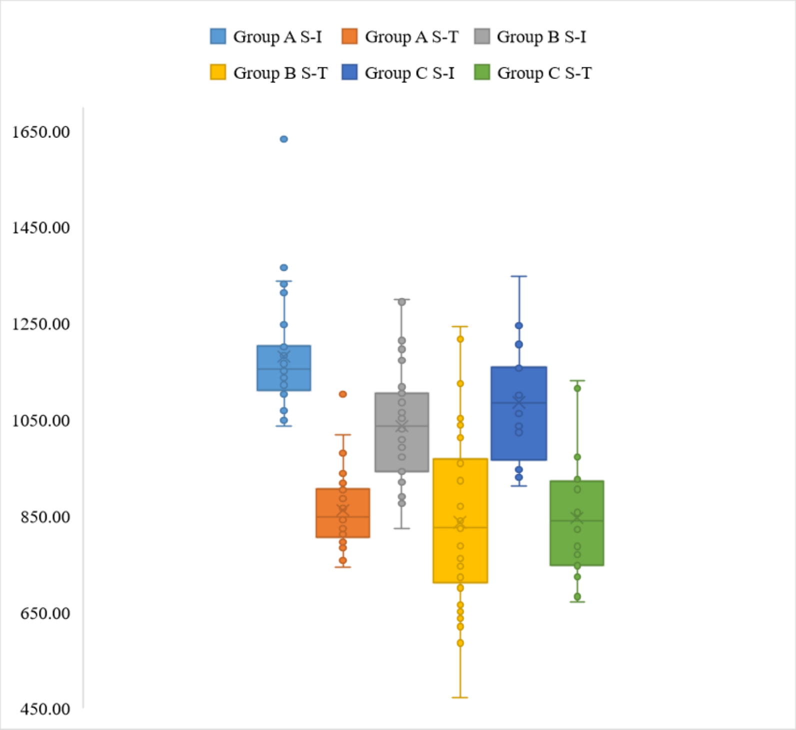


Figure 7.3a: Temporal EVS (in ms) of Group A, B, and C at S-I and S-T in RA-E

The average duration of the three groups' temporal EVS at S-I and S-T in the task RA-E are already presented in Section 7.1. In short, the group of British subjects had the shortest average temporal EVS at both the sentence initials and the sentence terminals, whereas the group of Chinese subjects had the longest average temporal EVS at both the sentence initials and the sentence terminals. More specifically, Group A's average temporal EVS at the sentence level is apparently longer than that of the subjects in both Group C and Group B. The differences between Group A and the other two groups are quite noticeable. However, *t*-tests are carried out to help determine whether these differences are statistically significant or whether they could reasonably occur randomly.

Table 7.3a presents the descriptive statistics and the inferential statistics from the *t*-tests carried out to compare Group A's temporal EVS with Group C's temporal EVS at the sentence initials and the sentence terminals respectively. Assuming that the Group A subjects have a significantly longer temporal EVS at both positions in the sentences, the *t*-tests are one-tailed and unpaired. Two-sample *t*-tests assuming either equal or unequal variances are used based on the variance of two datasets. Details of these *t*-tests are presented by a table in the appendices (see Appendix 4, Table 7.3a*).The Null Hypotheses are $\mu 1$ (Group A EVS at S-I) ≤ $\mu 2$ (Group C EVS at S-I) (a = 0.05), and $\mu 1$ (Group A EVS at S-T) ≤ $\mu 2$ (Group C EVS at S-T) (a = 0.05). To test the null hypotheses of zero difference, t-tests are carried out in conjunction with the Cohen's *d* effect size measurements to measure differences between the datasets.

Table 7.3a: Temporal EVS (in ms) of Group A and C at S-I and S-T in RA-E

RA-E	S-I		S-T	
Group	Group A	Group C	Group A	Group C
μ	1181.64	1085.36	861.62	846.35
SD	119.35	120.22	78.19	132.14
p (T⩽t) one tail	0.003806 < 0.01		0.304984< 0.5	
Cohen's *d*	0.8037 > 0.8		0.14 < 0.2	
*H_0	µ1 (Group A EVS)⩽µ2 (Group C EVS) (a = 0.05)			

The differences between temporal EVS of Group A and Group C are relatively noticeable at S-I. The *t*-statistic (2.786372) is considerably higher than the one-tail critical *t*-value (1.677224). In the one-tail two-sample *t*-test assuming equal variances, the *p*-value, which is much smaller than 0.01, shows that the average temporal EVS does represent the overall result: Group A's temporal EVS at S-I is longer than Group C's temporal EVS at S-I in RA-E. There is an extremely low probability that the differences between Group A's and Group C's temporal EVS at S-I occurred randomly. Hence, the present research has an extremely high confidence level to reject the H_0. However, things are different at S-T. The p-value is between 0.1 and 0.5. The effect size measures of the two *t*-tests are calculated. Cohen's $d = (M_2 - M_1) / SD_{\text{pooled}}$. For the first *t*-test assuming equal variances, Cohen's $d = 0.8037$; for the second *t*-test assuming equal variances, Cohen's $d = 0.1406$. The first measure exceeds the reference point of a *large* effect size, whereas the second measure is considered a very small effect size. Overall, nevertheless, since it was found earlier in Section 6.1 that the native German speakers in Group C invested a larger amount of cognitive effort during RA-E than the native Chinese speakers in Group A, this indicates that a longer temporal EVS at the sentence boundaries might signal a larger investment of cognitive effort.

English is the L2 of both Group A and Group C. Since neither group is reading a passage written in their L1, a further conclusion from these results might be that the length of one's temporal EVS when reading in one's L2 might have something to do with L2 competence. A comparable L2 competence between these two groups of speakers with different L1s was sought to be achieved when subjects were recruited based on their average age and English competence. In the following discussion, Group A's and Group B's temporal EVS results are investigated in the hope of finding a relationship between the length of one's temporal EVS and subjects' cognitive effort from another angle.

Table 7.3b presents the descriptive statistics and the inferential statistics from the *t*-tests carried out to compare Group A's temporal EVS with Group B's temporal EVS at the sentence initials and the sentence terminals respectively. Assuming that the Group A subjects have a significantly longer temporal EVS at both critical positions in the sentences, two-sample *t*-tests assuming either equal and unequal variances are used based on the variance of two datasets. The Null Hypotheses are $\mu 1$ (Group A EVS at S-I) $\leqslant$ $\mu 2$ (Group B EVS at S-I) (a = 0.05), and $\mu 1$ (Group A EVS at S-T) $\leqslant$ $\mu 2$ (Group B EVS at S-T) (a = 0.05). Details of these *t*-tests are presented in a table in the appendices (see Appendix 4, Table 7.3b*). To test the null hypotheses of zero difference between Group A and Group B, the *t*-tests are carried out in conjunction with the Cohen's *d* effect size measurements to measure differences between the datasets.

Table 7.3b: Temporal EVS (in ms) of Group A and B at S-I and S-T in RA-E

RA-E	S-I		S-T	
Group	Group A	Group B	Group A	Group B
μ	1181.64	1036.56	861.62	837.19
SD	119.35	113.65	78.19	180.94
p (T≤t) one tail	2.30927E-06 < 0.01		0.236346 < 0.5	
Cohen's *d*	1.25 > 0.8		0.18 < 0.2	
*H_0	$\mu 1$ (Group A EVS)$\leqslant\mu 2$ (Group B EVS) (a = 0.05)			

In the *t*-test, *t*-statistic (5.013553) is much higher than the one-tail critical *t*-value (1.669402). A *p*-value much lower than 0.01 is obtained. The present research has a very high confidence level that the two datasets (temporal EVS at S-I) are significantly different. Therefore, the H_0 is rejected and the researcher argues that Group A's temporal EVS is significantly and statistically longer than Group B's temporal EVS at sentence initials. On the other hand, a *p*-value between 0.1 and 0.5 is obtained from the one-tail two-sample *t*-tests assuming unequal variances. Also, the effect size measures of the two *t*-tests are calculated through the formula: Cohen's $d = (M_2 - M_1)/SD_{\text{pooled}}$. For the first *t*-test, Cohen's $d = 1.2450$; for the second *t*-test, Cohen's $d = 0.1753$. The first measure is very *large* effect size, whereas the second measure is not. Therefore, it is difficult to draw a conclusion from temporal EVS at S-T at this point. However, since it was suggested in Chapter 6 that the native Chinese speakers in Group A invested greater cognition during RA-E than the native English speakers in Group B. Again, this indicates that a longer temporal EVS at sentence initials might signal a larger amount of cognition investment.

Finally, it is noted that Group C's average temporal EVS at the sentence initials (S-I) and sentence terminals (S-T) is only marginally longer than that of Group B. To see if the difference of temporal EVS between Group B and Group C in RA-E is significant, two-sample *t*-tests assuming equal and unequal variances are used. Assuming that the Group C subjects have a significantly longer temporal EVS at both critical positions in the sentence, the Null Hypotheses are $\mu 1$ (Group C EVS at S-I) $\leq \mu 2$ (Group B EVS at S-I) ($a = 0.05$) and $\mu 1$ (Group C EVS at S-T) $\leq \mu 2$ (Group B EVS at S-T) ($a = 0.05$). Table 7.3c presents the descriptive statistics and the inferential statistics.

Table 7.3c: Temporal EVS (in ms) of Group B and C at S-I and S-T in RA-E

RA-E	S-I		S-T	
Group	Group C	Group B	Group C	Group B
μ	1085.36	1036.56	846.35	837.19
SD	120.22	113.65	132.14	180.94
p (T≤t) one tail	0.069723 < 0.1		0.421892 < 0.5	
Cohen's *d*	0.2 < 0.42 < 0.5		0.06 < 0.2	
*H_0	$\mu 1$ (EVS at S-I) $\leq \mu 2$ (EVS at S-T) ($a = 0.05$)			

Details of these *t*-tests are presented in a table in the appendices (see Appendix 4, Table 7.3c*). Neither *t*-statistic is higher than the one-tail critical *t*-value (1.6741). In light of the *p*-values ($p < 0.1$; $p < 0.5$), the present research does not have a high confidence level to entirely reject the H_0. The effect sizes of the two *t*-tests are measured respectively. The *d* values are 0.4172 and 0.0578. Neither measure is a *large* effect size. Although the results are acceptable, they do not irrefutably show that Group C's temporal EVS is significantly longer than Group B's temporal EVS. Although descriptively, Group C has a longer temporal EVS than Group B, the difference between their temporal EVS at either the sentence initials or the sentence terminals is not statistically significant. Since a hierarchy of their investment of cognitive effort was formed in Chapter 6 (Group A > Group C ≥ Group B), the fact that there is no significant difference between Group C's and Group B's temporal EVS in this reading task does not seem surprising. In fact, it demonstrates consistency between temporal EVS and other major indicators used in Chapter 6.

In summary, the initial hypothesis regarding the relationship between the length of temporal EVS and cognitive effort is reported in this section. Based on the hierarchy of cognitive effort formed in Chapter 6 (Group A > Group C ≥ Group B), it was found the average temporal EVS tends to be longer at both the sentence initials and the sentence terminals when the subject is devoting more effort to the process. The difference is particularly significant at sentence innitials, whereas it is less significant at sentence terminals.

7.3.2 Temporal EVS of Group A and C in RA and STR

In this section, the discussion continues with an intra-group examination within Group A and Group C, aiming at double checking the relationship found between the length of temporal EVS and cognitive effort in Section 7.3.1. As Macizo and Bajo (2004: 199) stated, "the main differences between reading for understanding and reading for translation were at the clause boundary, and these differences were independent of the direction of the translation". Hence, the sentence initials and the sentence terminals act as the clause boundaries under investigation in this section.

Similar to the comparison in Section 7.2, the intra-group comparison of the temporal EVS at the sentence level takes place within the test groups. Group A's temporal EVS in the four tasks (RA-C, RA-E, STR C-E, STR E-C) are contrasted with each other and Group C's temporal EVS results in RA-E are contrasted with those in STR E-G. In Chapter 6, three hierarchies of the cognitive effort investment were developed (Group A's cognitive effort investment in STR C-E > Group A's cognitive effort investment in RA-C, Group A's cognitive effort investment in STR E-C > Group A's cognitive effort investment in RA-E, Group C's cognitive effort investment in STR E-G > Group C's cognitive effort investment in RA-E) with the help of four eye-tracking indicators, namely pupil dilation (PCPD value), total fixation duration (TFD), fixation count (FC), and the sum of fixation and saccade duration (FSD). Hence, these hierarchies are used to establish the correlation between the temporal EVS length and cognitive effort in this section.

The first comparison is made between Group C's temporal EVS at the sentence level in the RA-E task and the STR E-G task, during which the subjects read the same passage written in English. The circles in Figure 7.3b represent the length of the average temporal EVS at the sentence initials and the sentence terminals during RA-E. The different shapes represent temporal EVS at either sentence initials or sentence terminals during each task.

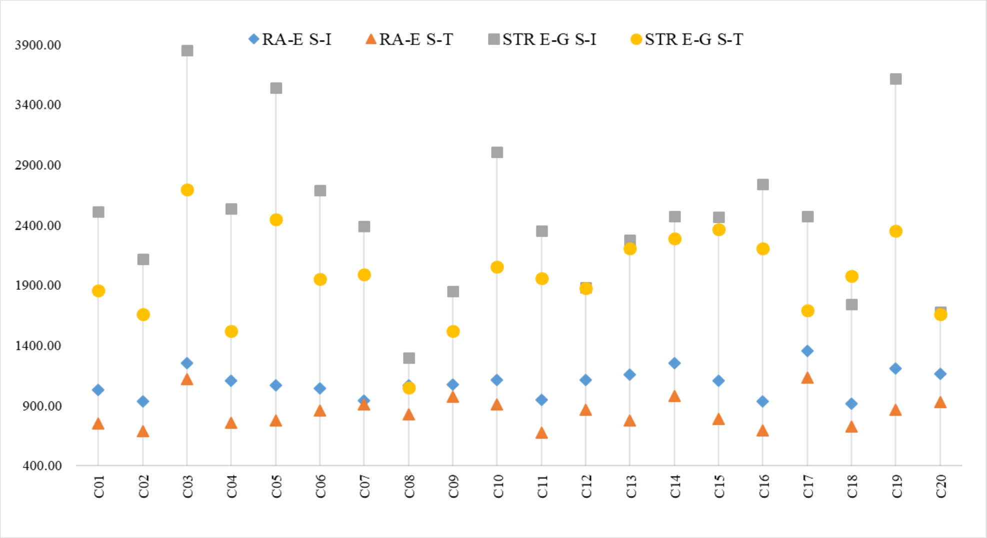


Figure 7.3b: Temporal EVS (in ms) of Group C at S-I and S-T in RA-E and STR E-G

The average length of Group C's temporal EVS at the sentence initials is 1085.36 ms (n = 20, *SD* = 120.22) in RA-E and 2471.14 ms (n = 20, *SD* = 658.86) in STR E-G. Their average temporal EVS at the sentence terminals is 846.35 ms (n = 20, *SD* = 132.14) in RA-E and 1961.69 ms (n = 20, *SD* = 385.76) in STR E-G. Some quantitative differences are shown in the graph. In comparison, the average temporal EVS is longer in the STR E-G task, at both the sentence initials and the sentence terminals. Two paired samples *t*-

tests are conducted to corroborate this hypothesis. Assuming that the Group C subjects have a significantly longer temporal EVS in the STR tasks at both the sentence initials and the sentence terminals, the *t*-test is one-tailed and paired. The Null Hypothesis is $\mu 1$ (EVS in STR) $\leq \mu 2$ (EVS in RA) (a = 0.05). Table 7.3d presents the descriptive and inferential results. Details of these *t*-tests are presented by a table in appendices (see Appendix 4, Table 7.3d*).

Table 7.3d: Temporal EVS (in ms) of Group C at S-I and S-T in RA-E and STR E-G

Group C	S-I		S-T	
Task	STR E-G	RA-E	STR E-G	RA-E
μ	2471.14	1085.36	1961.69	846.35
SD	658.86	120.22	385.76	132.14
p (T⩽t) one tail	3.62578E-09 < 0.1		5.35189E-11 < 0.1	
Cohen's *d*	2.93 > 0.8		3.87 > 0.8	
*H_0	µ1 (EVS in STR)⩽µ2 (EVS in RA) (a = 0.05)			

Both *t*-statistics are considerably higher than the one-tail critical *t*-values. The paired two-sample *t*-tests (one-tail) carried out with the dataset yield very low *p*-values ($p < 0.05$), which reinforce the finding of the descriptive statistics: the average temporal EVS is longer in the STR E-G task and shorter in the RA-E task, at both the sentence initials and the sentence terminals. Meanwhile, the effect size of the two *t*-tests are calculated. In the first *t*-test, Cohen's $d = 2.9262$; in the second *t*-test, Cohen's $d = 3.8682$. Both results significantly exceed the reference point of a *large* effect size. Since a) it is found that German subjects in Group C invested more effort during sight translating the passage from English to German than reading the English text aloud (in Chapter 6), b) the duration of temporal EVS is found in line with a trio of major eye-tracking indicators regarding the investment of cognitive effort (in Section 7.2.1), it is suggested that a longer temporal EVS might have a positive correlation with a higher level of cognition investment in a reading-speaking task.

The second comparison is made between Group A's temporal EVS at the sentence level in the RA-E task and the STR E-C task, during which the subjects were reading the same passage written in English. The circles in Figure 7.3c represent the average temporal EVS at 10 sentence initials and 10 sentence terminals during RA-E and STR E-C.

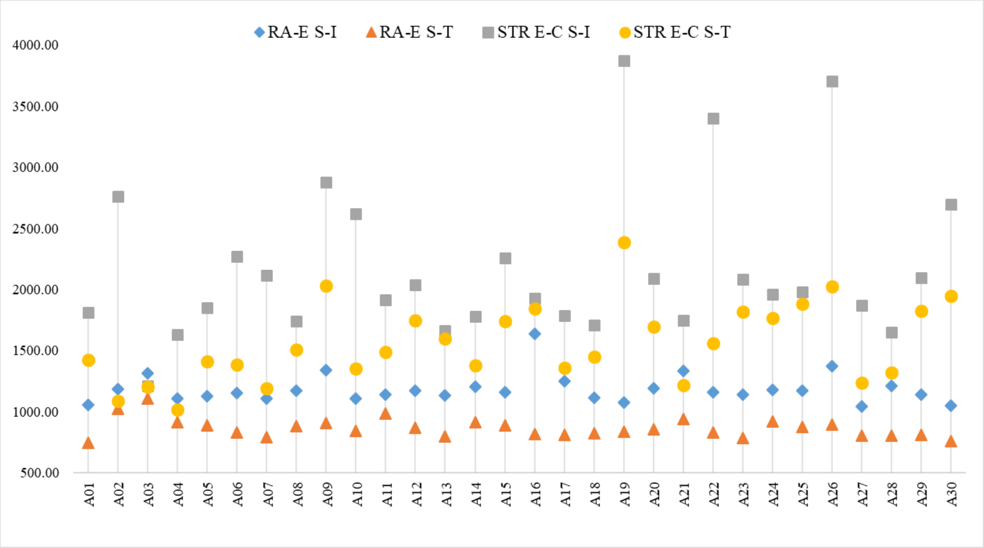


Figure 7.3c: Temporal EVS (in ms) of Group A at S-I and S-T in RA-E and STR E-C

Figure 7.3c shows the remarkable differences among the four datasets. At the sentence initials, Group A's average temporal EVS is 1181.64ms (n = 30, *SD* = 119.35) during RA-E and 2166.5 ms (n = 30, *SD* = 623.72) during STR E-C. At the sentence terminals, Group A's average temporal EVS is 861.62 ms (n = 30, *SD* = 78.19) during RA-E and 1559.32 ms (n = 30, *SD* = 320.64) during STR E-C. In contrast, the average length of subjects' temporal EVS in the STR E-C task is longer than that in the RA-E task, regardless of whether it is measured at the sentence initials or the sentence terminals. Since sight translating from English to Chinese consumes more cognitive effort than reading English aloud, this finding coincides with the initial hypothesis: the duration of temporal EVS has a positive relation with the intensity of cognitive processing. Meanwhile, this finding seems to be in line with what is found in the first comparison between Group C's temporal EVS in RA-E and STR E-G. To gain support for the initial hypothesis, the following *t*-tests are carried out. Assuming that Group A subjects have a significantly longer temporal EVS in the STR E-C task than in the RA-E task at the two critical positions in the sentences, the *t*-test is one-tailed and paired. The Null Hypothesis is $\mu 1$ (EVS in STR E-C) $\leqslant$ $\mu 2$ (EVS in RA-E) (a = 0.05). Both the descriptive statistics and the inferential statistics are presented in Table 7.3e. Details of these *t*-tests are presented by a table in appendices (see Appendix 4, Table 7.3e*).

Table 7.3e: Temporal EVS (in ms) of Group A at S-I and S-T in RA-E and STR E-C

Group A	S-I		S-T	
Task	STR E-C	RA-E	STR E-C	RA-E
μ	2166.50	1181.64	1559.32	861.62
SD	623.72	119.35	320.64	78.19
p ($T \leqslant t$) one tail	1.26714E-09 < 0.1		4.14199E-12 < 0.1	
Cohen's *d*	2.19 > 0.8		2.99 > 0.8	
*H_0	$\mu 1$ (EVS in STR) $\leqslant$ $\mu 2$ (EVS in RA) (a = 0.05)			

The two *t*-statistics are considerably higher than the one-tail critical *t*-values. The paired two-sample *t*-tests (one-tail) carried out with the dataset yield very low *p*-values. The significant difference between temporal EVS during RA and temporal EVS during STR at the sentence initials and the sentence terminals are both found at $p < 0.01$ level. Therefore, I can confidently say that the average length of these 20 German subjects' temporal EVS is longer in the STR E-G task and shorter in the RA-E task, regardless of whether it is measured at the sentence initials or the sentence terminals. Meanwhile, the effect size of the two *t*-tests were calculated. In the first *t*-test, Cohen's d = 2.1933; in the second *t*-test, Cohen's d = 2.9897. Both results suggest that the effect sizes are large, as 0.8 is considered as a *large* effect size. Therefore, the present research suggests that a longer temporal EVS symbolises a larger investment of cognitive effort in a reading-speaking task based on a) it is found that Chinese subjects in Group A invested more effort during sight translating the passage from English to Chinese than reading the English text aloud (in Chapter 6), b) the duration of temporal EVS is found in line with a trio of major eye-tracking indicators regarding the investment of cognitive effort (in Section 7.2.1). This interpretation, in fact, reinforces the hypothesis developed based on Group C's dataset.

The first and the second comparisons are made based on the two test groups' performance in the experiment using an English passage as ST. Since the Chinese students also completed another pair of RA and STR tasks with the same reading material written in Chinese, it provided an opportunity to test the hypothesis with another written language. The third comparison, therefore, is made between Group A's temporal EVS at the sentence level in the RA-C task and the STR C-E task, during which the subjects read the same passage written in Chinese. The circles in Figure 7.3d show the length of individuals' average temporal EVS at 12 sentence initials and 12 sentence terminals during RA-C and STR C-E.

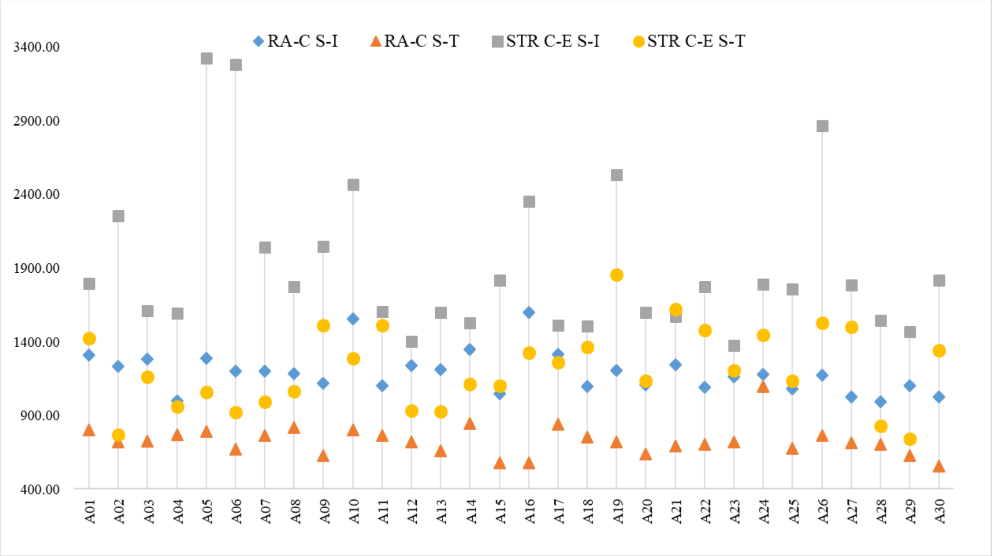


Figure 7.3d: Temporal EVS (in ms) of Group A at S-I and S-T in RA-C and STR C-E

Similar with the previous findings, the average length of the temporal EVS in the STR C-E task is longer than that in the RA-C task, at both the sentence initials and the sentence terminals. At the sentence initials, Group A's average temporal EVS is 1187.14 ms (n = 30, *SD* = 142.11) during RA-C; 1909.12 ms (n = 30, *SD* = 521.52) during STR C-E. At the sentence terminals, Group A's average temporal EVS is 724.70 ms (n = 30, *SD* = 103.28) during RA-C;

1213.69 ms (n = 30, *SD* = 273.67) during STR C-E.

Subsequently, two paired sample t-tests are conducted to corroborate this hypothesis. Assuming that Group A subjects have a significantly longer temporal EVS during the STR C-E task than during the RA-C task at the two critical positions in the sentences, the *t*-test is one-tailed and paired. The Null Hypothesis is $\mu 1$ (EVS in STR C-E) $\leq \mu 2$ (EVS in RA-C) (a = 0.05). Both the descriptive statistics and inferential statistics are presented in Table 7.3f. Details of these *t*-tests are presented in a table in the appendices (see Appendix 4, Table 7.3f*).

Table 7.3f: Temporal EVS (in ms) of Group A at S-I and S-T in RA-C and STR C-E

Group A	S-I		S-T	
Task	STR C-E	RA-C	STR C-E	RA-C
μ	1909.12	1187.14	1213.69	724.70
SD	521.52	142.11	273.67	103.28
p (T≤t) one tail	4.02516E-09 < 0.1		8.30437E-11 < 0.1	
Cohen's *d*	1.89 > 0.8		2.36 > 0.8	
*H_0	$\mu 1$ (EVS in STR)$\leq \mu 2$ (EVS in RA) (a = 0.05)			

The *t*-statistics (7.9981; 9.5978) are much higher than the one-tail critical *t*-values (1.6991; 1.6991). The paired two-sample *t*-tests (one-tail) carried out with the dataset yield very low *p*-values ($p < 0.01$), which reinforce the initial hypothesis once again. Also, the effect size of the two *t*-tests were calculated. The two *d* values are 1.8889 and 2.3641. Both effect sizes exceed the reference point of a *large* effect size. Therefore, the present research suggests that a longer temporal EVS symbolises a larger investment of cognitive effort in a reading-speaking task based on a) it is found that Chinese subjects in Group A invested more effort during sight translating the passage from Chinese to English than reading the Chinese text aloud (in Chapter 6), b) the duration of temporal EVS is found in line with a trio of major eye-tracking indicators regarding the investment of cognitive effort (in Section 7.2.1). This interpretation, reinforces the hypothesis that since sight translating the passage from Chinese to English is much more effort-consuming than reading the Chinese text aloud, a longer temporal EVS in a reading-speaking task and a higher level of cognition investment correspond with each other.

In this study, employing temporal EVS as an indicator of cognitive effort is presented as a viable approach for RA and STR process research. In fact, a similar suggestion was made about written translation process research. Dragsted and Hansen (2008) selected three words in the middle of a ST to calculate their subjects' eye-key spans. They found that the eye-key span is a common phenomenon during translation and its duration associates positively with the amount of cognitive investment. Although the present study observed remarkable interindividual differences in the length of temporal EVS during RA and STR, the data is indicative of fundamental behavioural differences in the dynamics between the eyes and the voice in different reading modes. The analyses support the overall assumption that there is a positive correlation between the amount of cognitive effort invested in the reading-speaking task and the length of a person's temporal EVS at the sentence level. It confirms the use of temporal EVS as an indicator of subjects' investment of cognitive effort. Furthermore, it is shown that temporal EVS can lead researchers to the same conclusions with the help of a series of major eye-tracking indicators.

7.4 Temporal EVS during STR

After establishing the relationship between the length of temporal EVS and cognitive effort, this section (7.4) aims to apply temporal EVS as an indicator of cognitive investment in more analyses. The reliability and validity of temporal EVS is examined further by:

1) contrasting Group A's temporal EVS in the forward STR task (STR C-E) with Group A's temporal EVS in the backward STR task (STR E-C);
2) comparing Group A's and Group C's temporal EVS rate of increase from RA to sight translating (from RA-E to STR E-C and from RA-

E to STR E-G).

7.4.1 Temporal EVS in Forward and Backward STR

When it comes to studying translation directionality, the comparability of the STs is crucial. It is extremely challenging to compare the forward and backward STR tasks in this study because the two STR tasks use two different STs, written in either English or Chinese. In addition, the Chinese language is a logographic script with no blank spaces between the characters, which would impact the pattern of fixations and saccades. As suggested by Rayner (1998; 2009), factors such as the orthographic regularity, the complexity of the texts, how difficult the texts are to read, and the characteristics of the font, all have influences on the eye movement parameters. Hence, Krieber *et al.* (2017: 11) emphasised that such comparisons "need to be interpreted with caution, in particular when differences in difficulty and complexity of a text may have an influence". Due to the typographic differences between the two writing systems, the present study defines the differences between the two tasks using the temporal EVS parameter, the calculation of which does not depend on the writing system of the languages. In addition, to ensure comparability, meaning units under investigation in both tasks are sentences, which are natural segments in both languages. Comparability of the Chinese and English STs employed in the present study is ensured by choosing the same source material, which makes sure the two passages have the same register, style, and speed of delivery. Hence, in this section, the comparisons between temporal EVS in the STR E-C task and the STR C-E task are carried out. Figure 7.4a presents Group A's temporal EVS at the sentence level in the forward and backward STR tasks. The descriptive statistics are presented in Table 7.4a.

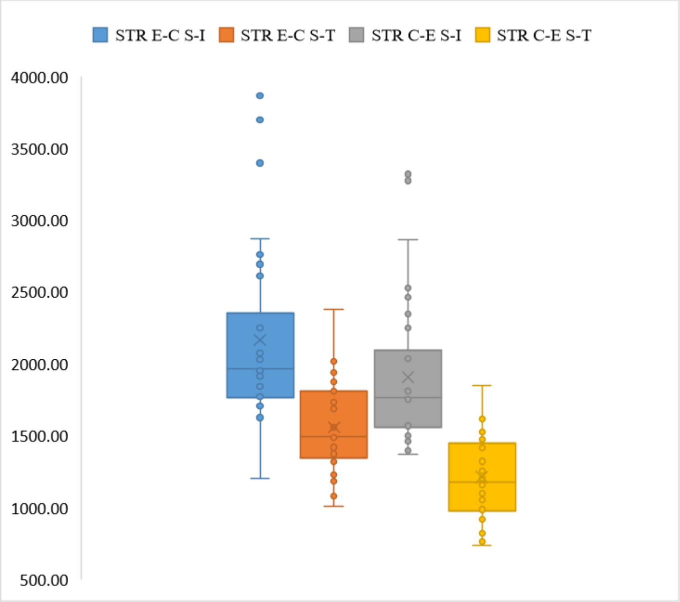


Figure 7.4a: Temporal EVS (in ms) of Group A at S-I and S-T in STR C-E and STR E-C

The average length of Group A's temporal EVS is 1909.12 ms (n = 30, *SD* = 521.52) at the sentence initials and 1213.69 ms (n = 30, *SD* = 273.67) at the sentence terminals in the task STR C-E. The average length of Group A's temporal EVS is 2166.5 ms (n = 30, *SD* = 623.72) at the sentence initials and 1559.32 ms (n = 30, *SD* = 320.64) at the sentence terminals in the task STR E-C. The *SD* values of temporal EVS in the STR task are generally high due to a relatively large interpersonal difference among subjects during STR. Nevertheless, all the standard values here are less than 30% of their corresponding mean values.

It appears that there are some differences between the temporal EVS during STR C-E and STR E-C. To assess whether the difference is statistically significant, *t*-tests are carried out, with some of the results presented in Table 7.4a. A more detailed presentation of these *t*-tests is in a table in the appendices (see Appendix 4, Table 7.4a*). Based on the averages, it is assumed that subjects in Group A have a significantly longer temporal EVS at the sentence level during STR E-C than during STR C-E. The Null Hypothesis is $\mu 1$ (STR E-C) $\leq$ $\mu 2$ (STR C-E) (a = 0.05).

Table 7.4a: Temporal EVS (in ms) of Group A at S-I and S-T in STR E-C and STR C-E

Group A	S-I		S-T	
Position	STR E-C	STR C-E	STR E-C	STR C-E
μ	2166.50	1909.12	1559.32	1213.69
SD	623.72	521.52	320.64	273.67
p (T≤t) one tail	0.013338 < 0.05		1.4912E-06 < 0.01	
Cohen's *d*	0.2< 0.45 < 0.5		1.16 > 0.8	
*H_0	µ1 (EVS in STR E-C)≤µ2 (EVS in STR C-E) (a = 0.05)			

The *t*-statistics (2.3349; 5.7720) turn out to be higher than the one-tail critical *t*-values (1.6991; 1.6991). The *p*-value ($p < 0.05$) of the first one-tail paired *t*-test conducted with Group A's temporal EVS at the sentence initials in the two STR tasks suggest that Group A's temporal EVS at the sentence initials is statistically longer when the subjects sight translated the English ST into Mandarin Chinese than when they sight translated the Chinese ST into English. A similar conclusion can also be drawn from the length of their temporal EVS at the sentence terminals. Furthermore, the effect size of the two *t*-tests were calculated [Cohen's $d = (M_2 - M_1) / SD_{\text{pooled}}$]. For the first *t*-test conducted with temporal EVS at S-I, the *d* value (0.4477) is considered as a *small* effect size, but is close to a *medium* effect size. For the *t*-test conducted with temporal EVS at S-T, the *d* value (1.1595) is indicative of a *large* effect size. The present study, therefore, suspects the difference between the mental processing during STR E-C and STR C-E lie at both the sentence initials and the sentence terminals.

In conclusion, based on temporal EVS, the present study suggests that the amount of cognitive effort devoted to dividing one's attention to reading ahead and speaking aloud during STR E-C is generally larger than that during STR C-E. This could be due to the translation directionality. For subjects in Group A, the difficulty of STR E-C lies mainly in processing the ST. It is less effort-consuming to process the input during STR C-E because the ST is written in their L1. In other words, it is more effort-consuming for these subjects to process the ST written in their L2 during STR, hence the temporal EVS in the STR E-C task is longer. This tentative conclusion, however, is not suggesting that all backward STR tasks require more effort than forward STR tasks. As an indicator, temporal EVS only shows us that the STR E-C task in this study consumes more cognitive effort than the STR C-E task. The fact that temporal EVS is able to reflect this demonstrates its value as an indicator of cognitive processing.

Furthermore, there is another further comparative possibility: the percentage increase of temporal EVS (the temporal EVS rate of increase) can be analysed to gain a further insight on the differences between Group A’s temporal EVS in the forward and backward STR tasks. This means examining temporal EVS rate of increase from RA-E to STR E-C and temporal EVS rate of increase from RA-C to STR C-E rather than contrasting the temporal EVS during STR E-C and STR C-E directly. With the individual’s temporal EVS in the RA tasks as the baseline, the increase of the length of temporal EVS in the STR tasks becomes an indicator that “truly refer[s] to the translation as opposed to the quality of language use” (Bayer-Hohenwarter 2010: 85). Temporal EVS rate of increase, in this manner, denotes the degree of the increasing demand of the cognitive resources in the STR tasks.

7.4.2 Temporal EVS in STR from English into Different Languages

After using temporal EVS as an indicator of cognitive investment in forward and backward STR tasks in Section 7.4.1, this section (7.4.2) revisits the cross-group comparison between the STR tasks from the same ST written in English to two different target languages: Chinese and German. The two STR tasks, STR E-C and STR E-G, are compared based on the fact that the two groups of subjects are sight translating from the same foreign language into their L1s. The goal is to search for differences between the Chinese and German subjects’ temporal EVS during STR L2-L1 and to re-examine the reliability and validity of temporal EVS as an indicator. Hence, this section starts with an analysis of the rate at which temporal EVS increases in the STR E-C task and the STR E-G task, then turns to apply major fixation-based measurements for evidence and finally compares the findings from temporal EVS and the other eye-tracking indicators.

Unlike the intra-group comparisons in the previous sections, where each subjects' dataset obtained in one task was contrasted with his/her own dataset obtained from another task, the comparison between the cognitive processing during STR E-C and STR E-G is based on two different groups' performance. When the two groups' performances in the two STR tasks are compared directly with each other the following questions may arise. Assuming that one subject from each group took 50 seconds to sight translate the text into their L1, does it mean the two subjects handled the STR task by investing the same amount of cognitive resources? What if by contrast, one of them only spent 25 seconds to read the English text aloud, whereas the other spent 50 seconds? Can researchers still draw the same conclusion as before? As a result, if the two STR tasks are compared directly, it might reveal many other variables that could bias the conclusion. The RA-E task is, therefore, used as the baseline task in this section and the percentage increase of each indicator is analysed.

The analysis starts with a comparison between the rates of increase in temporal EVS of Group A and C from RA-E to STR E-C and STR E-G. The temporal EVS rate of increase, in this manner, denotes the degree of increased cognitive demand. Descriptively, Chinese subjects' average temporal EVS increases from 1181.64 ms in the RA-E task to 2166.50 ms in the STR E-C task at the sentence initials, and from 861.62 ms in the RA-E task to 1559.32 ms in the STR E-C task at the sentence terminals. Meanwhile, the German subjects' average temporal EVS increases from 1085.36 ms in the RA-E task to 2471.14 ms in the STR E-G task at the sentence initials, and from 846.36 ms in the RA-E task to 1961.69 ms in the STR E-G task at the sentence terminals.

By subtracting the individual's temporal EVS in the RA task from their temporal EVS in the corresponding STR task, the increase of its length is obtained. Then, the percentage increase is calculated by dividing the initial temporal EVS in the RA task by the increase. Figure 7.4b presents the results at both the sentence initials and the sentence terminals.

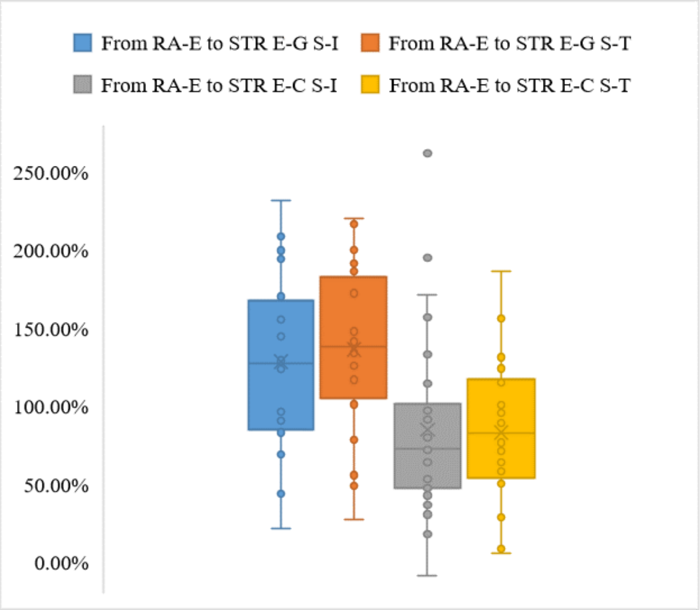


Figure 7.4b: Temporal EVS Rates of Increase (in Percentages) for Group A and C at S-I and S-T from RA-E to STR E-C and STR E-G

Compared with their own temporal EVS during RA-E, both Group A's and Group C's temporal EVS increase during the STR process. The rates of Group C's temporal EVS increase, both at the sentence initials and the sentence terminals, are higher than that of Group A. On average, Group A's temporal EVS increases by 85.04% (n = 30, *SD* = 0.5586) at the sentence initials and 83.02% (n = 30, *SD* = 0.4281) at the sentence terminals. Group C's temporal EVS increases by 128.38% (n = 20, *SD* = 0.5668) at the sentence initials and 136.33% (n = 20, *SD* = 0.5486) at the sentence terminals. It appears that the temporal EVS rate of increase of Group C is higher. Assuming that the German subjects have a significantly higher temporal EVS rate of increase at the sentence level, the *t*-tests should be one-tailed and unpaired. In addition, based on the variances, two paired two-sample *t*-tests assuming equal variances are carried out. The Null Hypothesis is $\mu 1$ (Group C temporal EVS rate of increase) ≤ $\mu 2$ (Group A temporal EVS rate of increase) (a = 0.05). The descriptive statistics and the inferential statistics are shown in Table 7.4b. Details of these *t*-tests are presented by a table in the appendices (see Appendix 4, Table 7.4b*).

Table 7.4b: Temporal EVS Rates of Increase (in Percentages) for Group A and C at S-I and S-T from RA-E to STR E-C and STR E-G

Temporal EVS Rate of Increase	S-I		S-T	
Task	STR E-G	STR E-C	STR E-G	STR E-C
μ	128.38	85.04	136.33	83.02
SD	56.68	55.86	54.86	42.81
p (T≤t) one tail	0.005137< 0.01		0.000174 < 0.01	
Cohen's *d*	0.5 < 0.77 < 0.8		1.08> 0.8	
*H_0	$\mu 1$ (STR E-G)≤$\mu 2$ (STR E-C) (a = 0.05)			

It turns out that the two *t*-statistics (2.6717; 3.8515) are both higher than the one-tail critical *t*-values (1.6772; 1.6772). The *p*-values in both *t*-tests are lower than 0.01. The high statistical power, therefore, allows us to reject the H_0, and conclude that this result is genuine with a high confidence level. In summary, Group C's temporal EVS rates of increase from RA-E to STR E-G are statistically higher than Group A's temporal EVS rates of increase from RA-E to STR E-C at both the sentence initials and the sentence terminals. Then, the effect size of the two *t*-tests were calculated respectively. For the first *t*-test conducted with temporal EVS rate of increase at S-I in the STR E-C task and the STR E-G task, the standardised effect size that conveys the size of an effect relative to the variability in the samples (Cohen's *d*) is 0.7702, which is very close to a *large* effect size. For the *t*-test conducted with temporal EVS rate of increase at S-T in the two STR tasks, the *d* value is 1.0834, which greatly exceeds a *large* effect size (the reference points for the *small*, the *medium*, and the *large* effects are 0.2, 0.5, and 0.8 respectively). To summarise, when the task changes from simply reading the text aloud to sight translating it into another language, Group C increases the amount of the cognitive resources invested in dividing their attention between reading and speaking during STR L2-L1 to a larger extent.

In addition, this section applies well-established eye-tracking measurements to test the conclusion drawn from temporal EVS rate of increase. If the same conclusion can be drawn by these indicators as well, it will demonstrate the reliability and validity of temporal EVS as an indicator of cognitive effort. Three well-established eye-tracking indicators, namely total fixation duration (TFD), fixation count (FC), and the sum of fixation and saccade duration (FSD) are used. The rates of increase in TFD, FC, and FSD from RA-E to STR E-C and STR E-G are presented in Figure 7.4c.

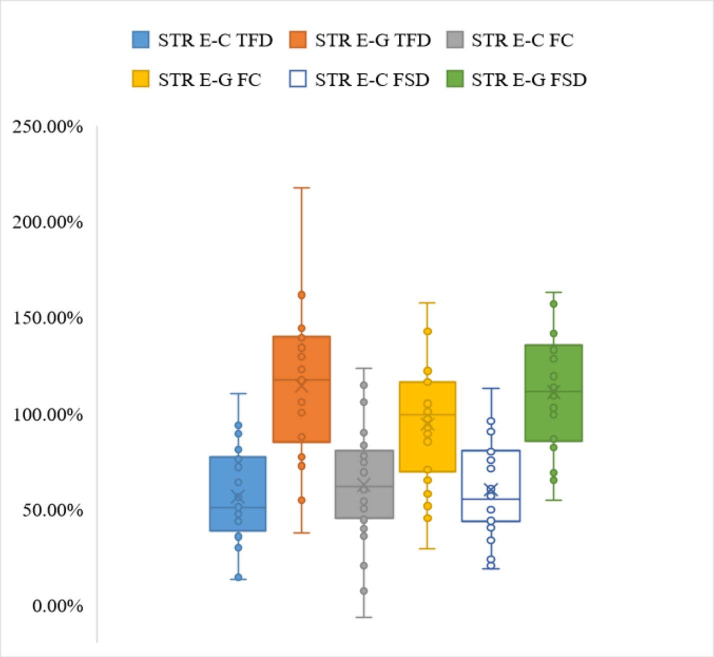


Figure 7.4c: TFD, FC, and FSD Rates of Increase (in Percentages) for Group A and C from RA-E to STR E-C and STR E-G

Compared with their TFD, FC, and FSD during RA-E, both Group A's and Group C's TFD, FC, and FSD increase during the STR process. From RA-E to STR E-C, Group A's average total fixation duration increases by 56.35% (n = 30, *SD* = 0.2570), their average fixation count increases by 62.55% (n = 30, *SD* = 0.2921), and their average FSD increases by 60.14% (n = 30, *SD* = 0.2510). From RA-E to STR E-G, Group C's average total fixation duration increases by 117.94% (n = 20, *SD* = 0.3766), their average fixation count increases by 97.16% (n = 20, *SD* = 0.2915), and their average FSD increases by 110.88% (n = 20, *SD* = 0.3090).

On average, the rates of increase in Group C's eye-tracking measurements are higher than the rates of increase for Group A. A series of *t*-tests are carried out. Since it is assumed that the German subjects have significantly higher rates of increase in TFD, FC, and FSD, the *t*-tests are both one-tailed and unpaired. The Null Hypothesis is $\mu 1$ (Group C rate of increase) $\leq \mu 2$ (Group A rate of increase) (a = 0.05). Three two-sample *t*-tests assuming equal variances are conducted. The descriptive statistics and the inferential statistics are shown in Table 7.4c. A more detailed presentation of these *t*-tests can be found in appendices (see Appendix 4, Table 7.3c*)

Table 7.4c: TFD, FC, and FSD Rates of Increase (in Percentages) for Group A and C from RA-E to STR E-C and STR E-G

Rate of Increase	TFD		FC		FSD	
Task	STR E-G	STR E-C	STR E-G	STR E-C	STR E-G	STR E-C
μ	117.94	56.35	97.16	62.55	110.88	60.14
SD	37.66	25.7	29.15	29.21	30.90	25.10
p (T≤t) one tail	5.5297E-09< 0.01		7.73491E-06 < 0.01		3.27134E-08 < 0.01	
Cohen's *d*	1.19 > 0.8		1.19 > 0.8		1.80 > 0.8	
*H_0	$\mu 1$ (STR E-G) $\leq \mu 2$ (STR E-C) (a = 0.05)					

Based on the variances, two-sample *t*-tests assuming equal variances are conducted. All of the three *t*-statistics (6.8844; 4.1079; 6.3812) are much higher than the one-tail critical *t*-values. The very low *p*-values ($p < 0.01$) reinforce the finding of the descriptive statistics. The high statistical power allows us to conclude the following with a high level of confidence: when the task changes from reading the text aloud to sight translating it into another language, the rates of increase in Group C's eye-tracking measurements are higher than those of Group A. Before giving a plausible explanation, the standardised effect sizes that convey the size of effects relative to the variability in the samples are measured. For the first *t*-test conducted with the rates of increase in TFD in STR E-C and STR C-E, the effect size of the *t*-

tests (Cohen's *d*) is 1.9102. For the *t*-test conducted with the rates of increase in FC in the two STR tasks, the Cohen's *d* is 1.1861. For the *t*-test conducted with the rates of increase in FSD in the two STR tasks, the standardised effect size (Cohen's *d*) is 1.8025. These measures all significantly exceed the reference point for a *large* effect (0.8). Hence, a plausible explanation is presented: the findings from TFD, FC, and FSD rates of increase indicate that Group C has increased the amount of cognitive investment from RA to STR to a larger extent than Group A.

To sum up, this section revisits the cross-group comparison between sight translating from the same ST written in English into two different target languages, Chinese and German. The RA-E task is used as the baseline task and various indicators' rates of increase from RA-E to STR E-C and STR E-G are analysed. The analysis of temporal EVS rate of increase of STR E-C and STR E-G is carried out first. Then, two traditional fixation-based measurements and one fixation-saccade-based indicator are applied to compare the findings from temporal EVS rates of increase, thus re-examining the reliability and validity of temporal EVS as an indicator of cognitive load.

A plausible explanation of these findings lies in the two groups' amount of cognitive invested in the same tasks. The fact that there is a statistically significant difference between the rates of increase in Group A's and Group C's temporal EVS in the STR E-C task and the STR E-G task indicates that the German subjects might have invested more cognitive resources than their Chinese counterparts. This conclusion, however, is not proposing that all the STR E-G tasks are more effort-consuming than the STR E-C tasks. As an indicator, temporal EVS shows us that the STR E-G task in this study was more demanding than the STR E-C task in terms of dividing one's attention and allocating it to both processing the written text and orally translating the text into another language. The fact that temporal EVS is able to reflect this demonstrates its strength as an indicator of cognitive processing. Since the results from analysing temporal EVS and a series of eye-tracking indicators are similar, the present study argues that temporal EVS is a valuable indicator of cognitive investment.

7.5 Summary

All the subjects' temporal EVS results are presented in Section 7.1. Tables display individuals' temporal EVS at sentence initials (S-I) and sentence terminals (S-T) in each task. Summary statistics including the mean and the standard deviation are also presented. It also touches upon the relationship between temporal EVS and other variables, including sentence length, reading speed, and spatial EVS.

Following data coding and presentation, the correlation coefficients between the length of temporal EVS and a series of eye-tracking measurements namely total fixation duration, fixation count, and the sum of fixation and saccade duration are analysed in Section 7.2.1. It is found that temporal EVS has strong correlation coefficients with individual eye-tracking measurements. The strong coefficients suggest that temporal EVS might be capable of indicating one's cognitive effort.

Subsequently, some comparisons are made in Section 7.2.2 to see how systematic changes in the independent variable "the location in sentences" affects the dependent variable "the duration of temporal EVS". It is shown that subjects tend to have longer temporal EVS at the sentence initials because the processing demand is the highest. At the sentence terminals, the subjects are preparing to finish the content in current sentence and prepare for the next sentence. Thus, a shorter temporal EVS lets the voice catch up with the eyes. This finding regarding temporal EVS is in line with Buswell's (1920) finding on spatial EVS, which stated that the subjects' average spatial EVS at the sentence initials is longer than that at the sentence terminals. Based on this marked correlation discovered between the duration of temporal EVS and the position in a sentence, the present study hypothesises that temporal EVS is not a random measure. It is potentially determined by the intensity of cognitive processing. Hence, the present research suggests that the length of temporal EVS has the potential to be used as an indicator of cognitive effort in eye-tracking based empirical studies.

After proving that temporal EVS does have the potential of being an indicator of cognitive investment, Section 7.3 tries to establish the relationship between the length of temporal EVS and the amount of cognitive effort invested in the reading-speaking tasks. Both intra-group and inter-group comparisons are made between tasks. Based on the hierarchies of cognitive effort discussed in Chapter 6, it is found that temporal EVS tends to be lengthened at both the sentence initials and the sentence terminals when a subject is devoting more effort to the process. Hence, when readers have comparable reading competence to the subjects in the present research, a longer temporal EVS symbolises a more substantial investment of cognitive effort in a reading-speaking task.

In the previous sections, the main finding shows that temporal EVS has a positive correlation with cognitive effort devoted to the tasks, meaning that temporal EVS can serve as a dynamic indicator for measuring cognitive effort. In Section 7.4.1, EVS is applied to compare STR E-C and STR C-E. The statistically significant difference between the temporal EVS at S-I and S-T in the two tasks indicates the difference between forward STR and backward STR. In Section 7.4.2, EVS was applied to compare STR E-C and STR E-G. The statistically significant difference found between the rates of increase in Group A’s and Group C’s temporal EVS in the STR L2-L1 tasks indicates that the language pair influences the amount of cognitive resources the subjects invested during STR. The fact that the EVS is able to indicate these differences between the STR processes demonstrates its advantage as an indicator of cognitive effort in the reading-speaking process.

Chapter 8. Conclusion

8.1 Answering the Research Questions

To summarise, the preliminary aim of this research was to apply well-established eye-tracking indicators, including pupil diameter, TTT, FSD, TFD, and FC, to investigate the hierarchies of cognitive investment in the RA and STR tasks carried out in the experiment. Two preliminary hypotheses regarding the hierarchies of cognitive resources invested during the process of RA and STR were made. The preliminary findings are:

1. In the RA-E task, the British subjects in Group B generally invested less cognitive effort than the other two groups in processing the reading material, reflected in the following hierarchy: Group A > Group C ≥ Group B;
2. Both the Chinese subjects in Group A and the German subjects in Group C invested more cognitive effort in the STR tasks than in the RA tasks, reflected in the following hierarchy: Group A STR C-E > Group A RA-C; Group A STR E-C > Group A RA-E; Group C STR E-G > Group C RA-E.

Based on the hierarchies formed in Chapter 6, the present study addresses the application of three eye-tracking indicators, namely MFD, FTS, and SPG. It is proposed that MFD might not be a reliable indicator of cognitive effort, whereas FTS and SPG have the potential to be used as measurement of one's cognitive investment in future eye-tracking based studies. Subsequently, the two preliminary findings regarding the hierarchies of cognitive resources invested during RA and STR are used as the basis for analysing temporal EVS, which reflects the dynamics between the eyes and voice (Laubrock and Kliegl 2015). In the present study, the hypothesis regarding temporal EVS is that there is a relationship between the length of temporal EVS and the amount of cognitive effort devoted during RA and STR. Four research questions were answered in Chapter 7.

Firstly, the present study aims to consider the potential of using temporal EVS as a latency measure in empirical studies and begins to analyse the nature of EVS by examining its correlation coefficient with individual eye-tracking measurements, including total fixation duration (TFD), fixation count (FC), and the sum of fixation and saccadic duration (FSD).

Although TFD, FC, and FSD are entirely based on eye-tracking data, temporal EVS does not depend solely on eye-movements. In brief, temporal EVS is not a conventional eye-movement based indicator. We should not only focus on the *E* (*eye*) of temporal EVS, but also look at the role of the *V* (*voice*) of temporal EVS because "articulatory output of a word presumably tells us that it no longer needs to be buffered in working memory" (Laubrock and Kliegl 2015: 2). Hence, the present research proposes that temporal EVS might correspond primarily to the subjects' ability to coordinate the input processing and the output execution. Nevertheless, strong correlation coefficients between the length of temporal EVS and the above-mentioned eye-tracking measurements are found. This showcases the potential relationship between the length of one's temporal EVS and cognitive investment.

The aforementioned eye-tracking measurements are macro-level indicators whereas temporal EVS is a micro-level measurement. While TFD, FC, and FSD indicate the depth or the intensity of cognitive processing, temporal EVS indicates eye-voice coordination. Because of the differences between temporal EVS and other eye-tracking indicators, the present study suggests that temporal EVS and those eye-tracking measurements might be representative of different sub-categories of cognitive effort devoted to completing a RA/STR task.

Secondly, temporal EVS in both RA and STR processes were analysed. During RA and STR, reading process typically follows a sequential scan to ensure a smooth and uninterrupted oral production. In general, the mind integrates all the information collected from successive fixations and forms a stable and coherent oral representation of the text during both RA and STR. When temporal EVS is applied to analyse RA and STR in the present study, it yields valuable results at the sentence initials and sentence terminals. Like its spatial counterpart, temporal EVS at sentence initials is significantly longer than temporal EVS at sentence terminals regardless of the task undertaken. The major difference between the length of temporal EVS at sentence initials and sentence terminals shows its high degree of adaptability and flexibility. This finding becomes the foundation for applying temporal EVS as an indicator of cognitive processing.

During reading-speaking, both the visual and cognitive sources of information are guiding the eyes (McConkie 1983: 75). The first guidance comes from the visual information gathered primarily from the periphery. Useful information, such as the length of the emerging words and punctuation, could be obtained within the perceptual span to the right of a fixation (Rayner 1983: 97). Hence, when a comma becomes visible in one's peripheral vision, the likelihood of it being recognised as the sentence boundary increases. The second indication comes from one's knowledge about the language and the text. The reader can guess that the sentence might end soon based on its content and grammatical structure. In this way, the two types of guidance work together to help coordinate the eyes and voice, although we do not know which one of them dominates in every case. Nevertheless, the present study argues that both the visual and cognitive sources of information act as a basis for the eye movement control at the sentence terminals and initials, and that they become joint forces to facilitate the maintenance of a temporal EVS during both RA and STR. The general patterns of temporal EVS within and across sentences' boundaries during both RA and STR show that thought determines the dynamics of EVS. Cognitive processing modulates the duration of temporal EVS. Therefore, the present study argues that the fact that the subjects have a flexible but not entirely random temporal EVS indicates that it is a manifestation of attention allocation and cognitive processing. Together with evidence found in the correlation coefficients between temporal EVS and major eye-tracking indicators, the present research suggests that temporal EVS has the potential to be used as a measurement for cognitive effort in eye-tracking based studies.

In addition, although both spatial EVS and temporal EVS showcase the difference in their length at S-I and S-T, they are manifestations of different reader traits and the reading process in the perspectives of the present study. McConkie (1983: 68) proposed two factors that are involved in eye guidance: the *spatial decision* and the *temporal decision*. One determines the locations of fixations and the other decides their durations. When an EVS, either spatial or temporal, is taking place during the RA or the STR process, we should consider which way these two decisions should be conceptualised. They could belong to a single decision: the eyes jump to a place further in the text and, at the same time, the pace at which the voice moves to this point is determined. In this case, the mind of the reader/translator acts as “a timing device, attempting to make an optimal estimate of how long the eyes should pause at each location” (McConkie 1983: 70). However, this may not be the case because the speed of one’s oral production, during either reading aloud or sight translation, is not fixed. Therefore, assuming that these two decisions do not take place at the same time, the present study hypothesises that the spatial decision is made first to launch the fixation to a point in the following text based on an estimation of the text difficulty and one’s peripheral vision. Then, the reader makes the temporal decision as to how long the eyes should fixate on this particular point and how quickly the voice takes to catch up. In this case, the mind is not working as a timing device that decides the temporal and spatial aspects of EVS at the onset of a forward fixation, but rather as an operator that continues to make constant changes to the decisions. Therefore, while spatial EVS is indicative of reading competence (e.g. Buswell 1920), temporal EVS can indicate cognitive load.

Thirdly, the relationship between one’s temporal EVS and cognitive effort invested in the RA and STR tasks is discussed based on the hierarchies of cognitive investment developed in Chapter 6. When the present study analyses different test groups' temporal EVS in the same task, it was found that the group who devoted the largest amount of cognitive effort displays the longest temporal EVS. This preliminary finding indicates that there might be a positive correlation between the length of EVS and the investment of cognitive resources. When the present study compares temporal EVS in the RA and STR tasks, temporal EVS at the sentence initials during RA is shorter than that during STR because the "lower level processes are usually enabled as soon as the word is encoded" (Carpenter and Just 1983: 304) during RA, whereas the STR task requires a higher-level integration. On the other hand, temporal EVS at the sentence terminals during STR is longer than that during RA due to the *end of sentence effects* in the STR tasks. Such effects cause extra processing burdens at the sentence terminals because readers need to finish the integration that was not completed earlier (Carpenter and Just 1983: 301). In short, temporal EVS has a positive correlation with the cognitive effort devoted to the tasks, meaning that EVS can measure the amount of cognitive effort when the readers have comparable reading competence to the subjects in the present research. In such cases, a longer temporal EVS represents a more substantial investment of cognitive effort in a reading-speaking task.

Having a temporal EVS during RA and STR coincides with what Carpenter and Just (1983: 303) called applying "a moving bin strategy". This binning process means "collecting input from several words before processing any one of them" (ibid.). If a bin is already full, either all or some of its contents are processed to make room for new information. From this perspective, the duration of temporal EVS varies depending on the readers' cognitive capacity that allows them to process more or less information while reading out loud at the same time. However, this does not mean that a longer EVS during oral reading is equivalent to a larger short-term memory buffer.

Instead, the shorter the temporal EVS, the more efficient the first-in-first-out short-term memory buffer is, which has a finite and limited capacity.

During reading-speaking, a reader integrates thoughts into the prior context held in the processing buffer as soon as he/she decodes a piece of information. The eyes, then, tend to proceed to the next fixation without waiting for the *bin* to be empty through oral production. In Rayner and Pollatsek's (1989: 181) words, "if the eyes moved further ahead, there would be a lot of undigested material" before the reader can produce it orally. When this happens, perhaps the "eyes need to wait for the voice because the size of the working memory buffer is limited" (Laubrock and Kliegl 2015: 17). However, if the cognitive processing and the voice output can catch up with the eyes, less information would remain undigested and the eyes do not need to slow down. For example, the available data from the cross-group comparison among Group A, B, and C shows that the faster readers have shorter temporal EVS in the RA-E task. Hence, if information processing and oral production could catch up with reading input, the *bin* becomes less full and the duration of temporal EVS in shortened. Therefore, a longer temporal EVS does not necessarily indicate a larger working memory capacity. In fact, it shows a fuller working memory buffer in some cases. The longer the information stays in the buffer due to the inability to vocalise it quickly, the more crowded the short-term memory buffer is. The fuller it is in the short-term memory buffer, the longer the temporal EVS. The longer the temporal EVS, the less coordinated the interplay between the eyes and the voice is.

Finally, the discussion focuses on what temporal EVS shows when it is applied to analyse different STR processes. This part of the discussion is specifically relevant to the STR tasks in the present study. The findings demonstrate the advantage of temporal EVS when it is applied to study different STR processes. When the present research applies temporal EVS to investigate different STR tasks, it shows that task type does have a significant influence on EVS patterns. The use of temporal EVS is examined by contrasting Group A’s temporal EVS in the forward STR task with that in the backward STR task and comparing temporal EVS in two STR tasks (STR E-C and STR E-G) carried out from English into two different TLs.

As a result, there is a statistically significant difference between the length of temporal EVS at the sentence initials and sentence terminals in the STR E-C task and STR C-E task. This finding indicates that the increase of cognitive effort between the forward STR and the backward STR process can be shown by analysing the subjects’ temporal EVS. However, that is not to say that all backward STR tasks require more effort than forward STR tasks. As an indicator, temporal EVS is showing us that the STR E-C task in this study is more demanding than the STR C-E task in terms of dividing the subjects’ attention between processing the written text and translating the text into another language simultaneously. By showing the difference between forward and backward STR, temporal EVS is shown to be very useful when the analysis is at a micro level. This finding suggests that if comparing forward and backward translation at a macro level does not yield clear results, perhaps an analysis of temporal EVS on a micro level would be beneficial. Furthermore, the fact that there is a statistically significant difference between the rates of increase of Group A’s and Group C’s temporal EVS in the STR L2-L1 tasks indicates that the German subjects have invested more cognitive resources than their Chinese counterparts. By contrasting the findings obtained from the temporal EVS rates of increase with the results from the eye-tracking indicators, the findings were in line with each other. Hence, the present study suggests that temporal EVS is a valuable indicator for studying the investment of the cognitive resources.

In addition, the fact that the RA-E task required less effort for Group C (see Chapter 6), while the STR L2-L1 task required more effort for Group C (see Chapter 7), suggests that what is indicated by the temporal EVS rate of increase in the STR tasks is not relevant to L2 competence, but sight translation competence. There could be many reasons why the German subjects, who are more competent readers of RA-E, devoted more cognitive effort during STR L2-L1. TL interference might be one of the reasons. For example, German, as a language is “characterised by a verbal-final position and extensive morphological structure of nouns” (Korpal 2012: 522), and this might have a significant impact on the subjects’ reading pattern. Following Seeber (2007), who explored the German verb-final structures during simultaneous interpreting, Korpal (2012) touched upon the topic *language-pair specificity* in his eye-tracking STR study and argued that the complexity of the STR task depends greatly on the particular language pair, especially when the ST is written in German. Korpal attributed this to the morphosyntactic characteristics of German, stating that it does have a bearing on the STR process. Therefore, this shows another potential characteristic of temporal EVS: as an indicator, it is sensitive to language interference during STR. Although it is beyond the scope of this study to investigate this matter further, it would be an interesting topic for future studies.

To sum up, the comparison between forward and backward STR and the comparison between STR from the same SL to different TLs both indicate that temporal EVS is able to reflect the influence of language pair on cognitive effort. These findings are particularly interesting because they provide researchers with a new way of studying STR directionality and comparing different STR processes in the future.

8.2 Contributions, Limitations, and Future Avenues of Research

In the past, some specific patterns of temporal EVS during RA were found, but the more global relationships between temporal EVS and cognition appear to be mostly absent. This study attempts to shed some light on the characteristics of temporal EVS during RA and STR. Temporal EVS during RA and STR reflects the subjects’ divided attention in several different sub-tasks: reading new information, processing the buffered meaning units, producing the processed units in the target language, monitoring the language output, and most importantly, synchronising the sub-tasks to keep the processing smooth. The length of temporal EVS is a visible indication representing the subjects' ability to synchronise the sub-tasks during RA and STR. The present study shows that temporal EVS is an indicator of the cognitive events that are taking place during both RA and STR. The present study argues that temporal EVS and spatial EVS should be used separately and should not be conflated. The temporal measurement (in milliseconds) calculated by eye-tracking is a more accurate measurement. Moreover, another contribution that the present eye-movements-based research aims to make to the field is showing creativity in finding ways to study the type of cognitive process we wish to understand.

This study, like many studies that have focused on temporal EVS during RA, relies on the data collected during the experiment to obtain evidence about the cognitive processing that occurs when temporal EVS takes place during the reading-speaking process. The present study uses the dataset as a window into the language processing mechanism to investigate whether temporal EVS could be an indicator of cognitive effort and test the hypotheses about the correlation between the length of temporal EVS and the amount of cognitive effort. The dataset allows the researcher to find a correlation between EVS and some individual eye-tracking measurements. It is also helpful for detecting the differences between various STR processes. Based on the data yielded by the experiment, as well as other published studies, it is the close relationship between the eyes and voice that allows the subjects' to successfully process the incoming information and convert it to an appropriate language output. As a result, this study argues that it is necessary to have strong mental faculties to coordinate incoming information and outgoing language output.

Since simultaneously executing the oculomotor and the articulatory movements reflects not only the visual sampling of the input but also the preparation of output (Laubrock and Kliegl 2015: 1), the present study argues that equal emphasis should be placed on both the *E (eye)* and the *V (voice)* when investigating temporal EVS during RA and STR. There is a connection between the eye and the notion of registration, and between the voice and the notion of utilisation. According to McConkie (1983: 87), *registration* in reading occurs "when the information becomes available to the brain" and *utilisation* in reading occurs "when the language processes are modified by the presence of that information". These two events coincide with the critical endpoints of temporal EVS. The *registration* defines the *E* (*eye)* in temporal EVS well as the onset of a fixation is when the transmission of the retinal encodings to the brain takes place. The *utilisation* represents the *V* (*voice*) in temporal EVS well as the start of an oral output indicates the effect of the language input, regardless of whether such a piece of information is fixated at this moment or not.

From the data collected during the experiment, the subjects' temporal EVS was never null or infinite during all the tasks. This led us to revisit McConkie's (1983: 70) *push view* and *pull view* of programming eye movements, bringing our attention to a different way of thinking about the nature of the eye movements. The length of temporal EVS, which primarily reflects the relationship between reading comprehension and oral production, seems to comply with *The Law of Action and Reaction*: when the voice starts pushing the eyes forwards, it receives a force pushing itself backwards; when the eyes are pulling the voice forwards, the eyes also receive a force pulling it backwards. Consequently, the reading speed and the length of temporal EVS are both restricted: the reading speed does not go too fast or too slow and the length of temporal EVS would not be reduced to zero or expand without a limit.

Due to the universal existence of temporal EVS, it seems that executing a temporal EVS during RA and STR is part of the routinised decisions (Jungermann, Pfister, and Fischer 2010). Moreover, the importance of maintaining an optimal time span between the eyes and the voice during reading-speaking has been emphasised by some researchers (e.g., Laubrock and Kliegl 2015). In one of the most recent studies on temporal EVS, Braze, Gong, and Nam (2018: 6) pointed out that “at a minimum, the eyes must lead the voice sufficiently that there is time to process visual input into speech motor commands before they are needed, but not by so much that memory available to process and store those commands is overtaxed”. The present study suggests that the length of one’s temporal EVS is affected by two abilities: the ability to lengthen temporal EVS and the ability to shorten temporal EVS. They play either a more significant or a minor role during reading-speaking. On the one hand, the ability to lengthen temporal EVS, depends largely on one’s speed of reading input. On the other hand, the ability to shorten EVS depends largely on the speed of speaking output. In this view, an experienced reader would develop a systematic system that exchanges the visual processing speed and the vocal processing speed to maintain a stable buffering mechanism.

Temporal EVS works as a buffering scheme that allows the continuous interpretive process to make use of the time lag between reading and speaking. The ability to process visual information and synchronise the input and output is essential. The length of temporal EVS will keep expanding as more visual unprocessed information accumulates. Since a better reader’s cognitive processing and oral production of the reading input takes place faster, the duration of his/her temporal EVS does not keep increasing because the segments processed successfully through oral output is released from the storage instantly. Hence, having the ability to shorten temporal EVS, rather than letting it expand uncontrollably, shows a high level of reading competence.

The fact that different computations may go on concurrently allows the reading and speaking processes to influence each other, not only by feeding the result of one to the other, but also by sharing the reader’s attention. Hence, as a metric that has been used to observe how readers read aloud, temporal

EVS demonstrates the reader's divided attention between reading and speaking. When the reader's attention is divided, the attention is split into comprehension and oral production (Zheng and Zhou 2018). Furthermore, it does not mean the processing time is divided in two halves for reading and speaking. Instead, it means one's attention is allocated to the two processes simultaneously. Nevertheless, having an EVS during RA and STR is not contradictory to the *eye-mind assumption*, as this assumption does not require one's cognitive system to consider only the words that are currently being fixated (Carpenter and Just 1983: 276).

Laubrock and Kliegl (2015: 17) suggest that temporal EVS varies in accordance with other cognitive, oculomotor, and articulatory demands during the task. Moreover, the task goal, style, and instruction are also important variables. Having a flexible temporal EVS shows that one has been updating his/her dynamic buffering system constantly in accordance with these variables. The results of the present study indicate that one possible component of achieving good online cognitive control is modifying the length of temporal EVS with respect to the task. Hence, developing a suitable temporal EVS during STR and the ability to adjust its duration when it is necessary could be helpful for novice translators to improve their STR skills.

STR is particularly interesting because this area is the one in which an eye-tracking methodology may result in progress towards understanding the cognitive process during translation. Also, STR may be a trend in the translation market and translator training in the foreseeable future. With the help of the speech recognition technologies, STR will hopefully "have a lot to offer in terms of saving time and effort without compromising the output quality significantly" (Dragsted and Hansen 2009: 602). Dragsted and Hansen (2009) argue that translators need to prepare for a change in their working routines due to the fast development of technologies. Hence, this study hopes to establish temporal EVS as an indicator of cognitive effort during reading-speaking processes to assist future researchers interested in investigating the

RA and STR processes with modern technology. Admittedly, using temporal EVS as an indicator in an empirical study is a currently time-consuming. However, the value of this latency measure should not be denied because it is difficult to calculate. Hopefully, in the near future, technicians will make the EVS calculation an automatic procedure by installing a speech recognition device within eye-tracking equipment. This speech recognition device would be better than an ordinary audio-recording device as it should be able to map out the specific time points automatically when a certain syllable is articulated, as well as recording which syllable is articulated at a specific time point.

Although some progress has been made in understanding temporal EVS based on empirical experiments, our understanding of temporal EVS could still benefit from some additional research angles and approaches. As mentioned in Section 1.5, triangulating another online data, such as the subjects' heart rate and blood pressure, or conducting a series of studies on temporal EVS and RA/STR products could compensate the limitations of the present study. Moreover, it would be worthwhile to verify the results of the present study by increasing the representative groups or the number of subjects. Furthermore, there are many more topics related to temporal EVS that are worth exploring in the future. For example, the relationship between time pressure and subjects' temporal EVS could be a fruitful study as one's reading behaviour is easily "affected by time pressure and the concern for smooth delivery" (Dragsted and Hansen 2009: 597). Finally, researchers, if possible, could carry out longitudinal studies to investigate the development of temporal EVS directly in future studies.

Appendices

Appendix 1

Material A

Page 1

The expansion of trade hasn't fully closed the gap between those of us who live on the cutting edge of the global economy and the billions around the world who live on the knife's edge of survival. This global gap requires more than compassion. It requires action. Global poverty is a powder keg that could be ignited by our indifference.

Page 2

In his first inaugural address, Thomas Jefferson warned of entangling alliances. But in our times, America cannot and must not disentangle itself from the world. If we want the world to embody our shared values, then we must assume a shared responsibility.

Page 3

We must embrace boldly and resolutely that duty to lead, to stand with our allies in word and deed, and to put a human face on the global economy so that expanded trade benefits all people in all nations, lifting lives and hopes all across the world.

Page 4

Third, we must remember that America cannot lead in the world unless here at home we weave the threads of our coat of many colors into the fabric of one America. As we become ever more diverse, we must work harder to unite around our common values and our common humanity.

Material B

Page 1

机会属于所有的美国公民；责任源自全体美国人民；所有美国人民组成了一个大家庭。我一直在为寻求一个更小、更现代化、更有效率、面对新时代的挑战充满创意和思想、永远把人民的利益放在第一位、永远面向未来的新型的美国政府而努力。

Page 2

我们要加倍努力地工作，克服生活中存在的种种分歧。于情于法，我们都要让我们的人民受到公正的待遇，不论他是哪一个民族、信仰何种宗教、什么性别或性倾向，或者何时来到这个国家。我们时时刻刻都要为了实现先辈们建立高度团结的美利坚合众国的梦想而奋斗。

Page 3

对我来说，当我离开总统宝座时，我充满更多的理想，比初进白宫时更加充满希望，并且坚信美国的好日子还在后面。我的总统任期就要结束了，但是我希望我为美国人民服务的日子永远不会结束。

Page 4
在我未来的岁月里，我再也不会担任一个能比美利坚合众国总统更高的职位、签订一个比美利坚合众国总统所能签署的更为神圣的契约了。当然，没有任何一个头衔能让我比作为一个美国公民更为自豪的了。谢谢你们！愿上帝保佑你们！愿上帝保佑美国！

Material C

Page 1
Tonight, I want to leave you with three thoughts about our future. First, America must maintain our record of fiscal responsibility. Through our last four budgets, we’ve turned record deficits to record surpluses, and we’ve been able to pay down $600 billion of our national debt, on track to be debt free by the end of the decade for the first time since 1835.

Page 2
Second, because the world is more connected every day in every way, America’s security and prosperity require us to continue to lead in the world. At this remarkable moment in history, more people live in freedom than ever before. Our alliances are stronger than ever. People all around the world look to America to be a force for peace and prosperity, freedom and security.

Material D

Page 1
作为总统，我所做的每一个决定，每一个行政命令，提议和签署的每一项法令，都在努力为美国人民提供工具和创造条件，去实现美国人民梦想的未来：一个美好的社会，繁荣的经济，清洁的环境，一个更自由、更安全、更繁荣的世界。

Page 2
这是一个极具变革的年代，你们为新的挑战做好了准备。是你们使我们的社会更强大，我们的家庭更健康和安全，我们的人民更富裕。同胞们，我们已迈进全球信息化的时代，这是美国复兴的伟大时代。

Appendix 2

Glossary

GLOSSARY

1. *Alliance* (n.)

A union or association formed for mutual benefit, especially between countries or organizations.

2. *Compassion* (n.)

Sympathetic pity and concern for the sufferings or misfortunes of others.

3. *Disentangle* (v.)

Free (something or someone) from something that they are entangled with.

4. *Entangle* (v.)

Cause to become twisted together with or caught in.

5. *Keg* (n.)

A small barrel, especially one of less than 10 gallons or (in the US) 30 gallons.

6. *Ignite* (v.)

Catch fire or cause to catch fire.

7. *Inaugural* (adj.)

Marking the beginning of an institution, activity, or period of office.

8. *Indifference* (n.)

Lack of interest, concern, or sympathy.

Appendix 3

Consent Form

CONSENT TO PARTICIPATE IN RESEARCH

Identification of Investigator & Purpose of Study

You are being invited to participate in a research study conducted by Hao Zhou, PhD student from School of Modern Languages and Cultures at Durham University. This study will contribute to the researcher's completion of her degree.

Research Procedures

This experiment includes a read-aloud task, a sight translation task and an interview.

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 generalizations about each group of examinees as a whole. All the data will be stored in a secure place accessible only to the researcher.

Participation and Withdrawal

Your participation is entirely voluntary. If you choose to participate, you can withdraw at any time without consequences of any kind.

Right as Research Subjects

You are not waiving any legal claims, rights or remedies because of your participation in this research study.

Questions about the Study

If you have questions or concerns about the study after its completion, please contact:
Researcher's Name: Hao Zhou
Department: School of Modern Languages and Cultures, Durham University
Email Address: hao.zhou@durham.ac.uk
Telephone: +0044 （0）7598470114

Giving of Consent

I have read this consent form and I understand what is being requested of me as a participant in this study. I freely consent to participate. I also agree to be recorded during the oral testing. I also give the researcher my consent for the use of my data for any anticipated future research.

Name of Participant: __________________(Signed) Date: __________________

Name of researcher: __________________(Signed) Date: __________________

Appendix 4 Tables Referenced in Chapter 6 and Chapter 7

Table 6.1.1a*: Two-sample T-test Assuming Equal Variances [TTT of Group A & B in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group A TTT)	*Variable 2 (Group B TTT)*
Mean	92.54866667	66.71428571
Variance	102.0422395	89.79566639
Observations	30	35
Pooled Variance	95.43297784	
Hypothesised Mean Difference	0	
df	63	
t Stat	10.6288626	
P(T<=t) one-tail	5.49287E-16	
t Critical one-tail	1.669402222	
P(T<=t) two-tail	1.09857E-15	
t Critical two-tail	1.998340543	

Table 6.1.1b*: Two-sample T-test Assuming Equal Variances [TTT of Group A & C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group A TTT)	*Variable 2 (Group C TTT)*
Mean	92.54866667	71.042
Variance	102.0422395	36.38326947
Observations	30	20
Pooled Variance	76.05223056	
Hypothesised Mean Difference	0	
df	48	
t Stat	8.542947727	
P(T<=t) one-tail	1.68551E-11	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	3.37103E-11	
t Critical two-tail	2.010634758	

Table 6.1.1c*: Two-sample T-test Assuming Equal Variances [TTT of Group B & C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group B TTT)	*Variable 2 (Group C TTT)*
Mean	71.042	66.71428571
Variance	36.38326947	89.79566639
Observations	20	35
Pooled Variance	70.64782598	
Hypothesised Mean Difference	0	
df	53	
t Stat	1.836863377	
P(T<=t) one-tail	0.035920108	
t Critical one-tail	1.674116237	
P(T<=t) two-tail	0.071840215	
t Critical two-tail	2.005745995	

Table 6.1.1d*: Two-sample T-test Assuming Equal Variances [FSD of Group A & B in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group A FSD)	*Variable 2 (Group B FSD)*
Mean	85.39775888	64.8864301
Variance	92.81884418	126.299828
Observations	30	35
Pooled Variance	110.8879466	
Hypothesised Mean Difference	0	
df	63	
t Stat	7.828699024	
P(T<=t) one-tail	3.55687E-11	
t Critical one-tail	1.669402222	
P(T<=t) two-tail	7.11374E-11	
t Critical two-tail	1.998340543	

Table 6.1.1e*: Two-sample T-test Assuming Equal Variances [FSD of Group A & C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group A FSD)	*Variable 2 (Group C FSD)*
Mean	85.39775888	69.11855885
Variance	92.81884418	52.21773834
Observations	30	20
Pooled Variance	76.74757312	
Hypothesised Mean Difference	0	
df	48	
t Stat	6.437116321	
P(T<=t) one-tail	2.68519E-08	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	5.37038E-08	
t Critical two-tail	2.010634758	

Table 6.1.1f*: Two-sample T-test Assuming Equal Variances [FSD of Group B & C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group C FSD)	*Variable 2 (Group B FSD)*
Mean	69.11855885	64.60071581
Variance	52.21773834	104.5772238
Observations	20	35
Pooled Variance	85.80684225	
Hypothesised Mean Difference	0	
df	53	
t Stat	1.739954348	
P(T<=t) one-tail	0.043835697	
t Critical one-tail	1.674116237	
P(T<=t) two-tail	0.087671394	
t Critical two-tail	2.005745995	

Table 6.1.1g*: Two-sample T-test Assuming Equal Variances [TFD of Group A & B in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group A TFD)	*Variable 2 (Group B TFD)*
Mean	76.422	55.94371429
Variance	96.37443724	83.4233005
Observations	30	35
Pooled Variance	89.38493488	
Hypothesised Mean Difference	0	
df	63	
t Stat	8.705613964	
P(T<=t) one-tail	1.04875E-12	
t Critical one-tail	1.669402222	
P(T<=t) two-tail	2.09749E-12	
t Critical two-tail	1.998340543	

Table 6.1.1h*: Two-sample T-test Assuming Equal Variances [TFD of Group A & C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group A TFD)	*Variable 2 (Group C TFD)*
Mean	76.422	58.973
Variance	96.37443724	35.34444316
Observations	30	20
Pooled Variance	72.21673125	
Hypothesised Mean Difference	0	
df	48	
t Stat	7.112827104	
P(T<=t) one-tail	2.46926E-09	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	4.93852E-09	
t Critical two-tail	2.010634758	

Table 6.1.1i*: Two-sample T-test Assuming Equal Variances [TFD of Group B & C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group C TFD)	*Variable 2 (Group B TFD)*
Mean	58.973	55.60085714
Variance	35.34444316	61.93785513
Observations	20	35
Pooled Variance	52.40436782	
Hypothesised Mean Difference	0	
df	53	
t Stat	1.661842636	
P(T<=t) one-tail	0.051224374	
t Critical one-tail	1.674116237	
P(T<=t) two-tail	0.102448747	
t Critical two-tail	2.005745995	

Table 6.1.1j*: Two-sample T-test Assuming Equal Variances [FC of Group A & B in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group A FC)	*Variable 2 (Group B FC)*
Mean	304.2333333	240.8857143
Variance	1043.84023	1358.751261
Observations	30	35
Pooled Variance	1213.792215	
Hypothesised Mean Difference	0	
df	63	
t Stat	7.307958859	
P(T<=t) one-tail	2.89813E-10	
t Critical one-tail	1.669402222	
P(T<=t) two-tail	5.79626E-10	
t Critical two-tail	1.998340543	

Table 6.1.1k*: Two-sample T-test Assuming Equal Variances [FC of Group A & C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group A FC)	*Variable 2 (Group C FC)*
Mean	304.2333333	256.35
Variance	1043.84023	513.5026316
Observations	30	20
Pooled Variance	833.9149306	
Hypothesised Mean Difference	0	
df	48	
t Stat	5.743995934	
P(T<=t) one-tail	3.07729E-07	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	6.15457E-07	
t Critical two-tail	2.010634758	

Table 6.1.1l*: Two-sample T-test Assuming Equal Variances [FC of Group B & C in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Variable 1 (Group C FC)	*Variable 2 (Group B FC)*
Mean	256.35	240.8857143
Variance	513.5026316	1358.751261
Observations	20	35
Pooled Variance	1055.737601	
Hypothesised Mean Difference	0	
df	53	
t Stat	1.697929076	
P(T<=t) one-tail	0.047692974	
t Critical one-tail	1.674116237	
P(T<=t) two-tail	0.095385949	
t Critical two-tail	2.005745995	

Table 6.1.2a*: Paired Two-sample T-test for Means [Pupil Size of Group C in RA-E & STR E-G]

T-Test: Paired Two Sample for Means

Group C Pupil Diameter	*Variable 1 (STR E-G)*	*Variable 2 (RA-E)*
Mean	3.336049781	3.241381083
Variance	0.149814684	0.118774278
Observations	20	20
Pearson Correlation	0.956388288	
Hypothesised Mean Difference	0	
df	19	
t Stat	3.652629319	
P(T<=t) one-tail	0.000846483	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	0.001692965	
t Critical two-tail	2.093024054	

Table 6.1.2b*: Paired Two-sample T-test for Means [Pupil Size of Group A in RA-C, STR C-E, RA-E, & STR E-C]

T-Test: Paired Two Sample for Means

Group A Pupil Diameter	*Variable 1 (STR E-C)*	*Variable 2 (RA-E)*
Mean	3.060560188	2.92037153
Variance	0.072079857	0.061074403
Observations	30	30
Pearson Correlation	0.958889884	
Hypothesised Mean Difference	0	
df	29	
t Stat	9.987311899	
P(T<=t) one-tail	3.39634E-11	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	6.79268E-11	
t Critical two-tail	2.045229642	

	Variable 1 (STR C-E)	*Variable 2 (RA-C)*
Mean	3.058910563	2.855247049
Variance	0.068672193	0.058740077
Observations	30	30
Pearson Correlation	0.967841458	
Hypothesised Mean Difference	0	
df	29	
t Stat	16.67985579	
P(T<=t) one-tail	1.05403E-16	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	2.10806E-16	
t Critical two-tail	2.045229642	

Table 6.1.2c*: Paired Two-sample T-test for Means [TTT of Group A & C in RA & STR]

T-Test: Paired Two Sample for Means

Group A TTT	*Variable 1 (STR C-E)*	*Variable 2 (RA-C)*
Mean	162.856	90.464
Variance	250.7097352	74.0300731
Observations	30	30
Pearson Correlation	0.073240058	
Hypothesised Mean Difference	0	
df	29	
t Stat	22.7119848	
P(T<=t) one-tail	2.55E-20	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	5.09363E-20	
t Critical two-tail	2.045229642	
Group A TTT	*Variable 1 (STR E-C)*	*Variable 2 (RA-E)*
Mean	156.258	92.54866667
Variance	464.7219683	102.0422395
Observations	30	30
Pearson Correlation	0.357730986	
Hypothesised Mean Difference	0	
df	29	
t Stat	17.21324569	
P(T<=t) one-tail	4.59E-17	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	9.17584E-17	
t Critical two-tail	2.045229642	
Group C TTT	*Variable 1 (STR E-G)*	*Variable 2 (RA-E)*
Mean	165.8745	71.042
Variance	300.7786997	36.38326947
Observations	20	20
Pearson Correlation	0.011738415	
Hypothesised Mean Difference	0	
df	19	
t Stat	23.18144027	
P(T<=t) one-tail	1.07096E-15	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	2.14192E-15	
t Critical two-tail	2.093024054	

Table 6.1.2d*: Paired Two-sample T-test for Means [FSD of Group A & C in RA & STR]

T-Test: Paired Two Sample for Means

Group A FSD	*Variable 1 (STR C-E)*	*Variable 2 (RA-C)*
Mean	138.4340611	83.17209213
Variance	214.6460584	117.1830362
Observations	30	30
Pearson Correlation	0.384033219	
Hypothesised Mean Difference	0	
df	29	
t Stat	20.88624641	
P(T<=t) one-tail	2.52659E-19	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	5.05318E-19	
t Critical two-tail	2.045229642	
Group A FSD	*Variable 1 (STR E-C)*	*Variable 2 (RA-E)*
Mean	135.621147	85.39775888
Variance	357.7460831	92.81884418
Observations	30	30
Pearson Correlation	0.254405991	
Hypothesised Mean Difference	0	
df	29	
t Stat	14.54178879	
P(T<=t) one-tail	3.72845E-15	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	7.45689E-15	
t Critical two-tail	2.045229642	
Group C FSD	*Variable 1 (STR E-G)*	*Variable 2 (RA-E)*
Mean	144.5650907	69.11855885
Variance	336.9951507	52.21773834
Observations	20	20
Pearson Correlation	0.210761128	
Hypothesised Mean Difference	0	
df	19	
t Stat	18.48157975	
P(T<=t) one-tail	6.66E-14	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	1.33117E-13	
t Critical two-tail	2.093024054	

Table 6.1.2e*: Paired Two-sample T-test for Means [TFD of Group A & C in RA & STR]

T-Test: Paired Two Sample for Means

Group A TFD	*Variable 1 (STR C-E)*	*Variable 2 (RA-C)*
Mean	121.7826667	74.27433333
Variance	232.8566892	77.9998392
Observations	30	30
Pearson Correlation	0.234057462	
Hypothesised Mean Difference	0	
df	29	
t Stat	16.53129828	
P(T<=t) one-tail	1.33398E-16	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	2.66795E-16	
t Critical two-tail	2.045229642	

Group A TFD	*Variable 1 (STR E-C)*	*Variable 2 (RA-E)*
Mean	118.3426667	76.422
Variance	322.7799168	96.37443724
Observations	30	30
Pearson Correlation	0.31042738	
Hypothesised Mean Difference	0	
df	29	
t Stat	13.04824174	
P(T<=t) one-tail	5.76835E-14	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	1.15367E-13	
t Critical two-tail	2.045229642	

Group C TFD	*Variable 1 (STR E-G)*	*Variable 2 (RA-E)*
Mean	127.3015	58.973
Variance	314.9009713	35.34444316
Observations	20	20
Pearson Correlation	0.040689612	
Hypothesised Mean Difference	0	
df	19	
t Stat	16.53178751	
P(T<=t) one-tail	4.91784E-13	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	9.83568E-13	
t Critical two-tail	2.093024054	

Table 6.1.2f*: Paired Two-sample T-test for Means [FC of Group A & C in RA & STR]

T-Test: Paired Two Sample for Means

Group A FC	*Variable 1 (STR C-E)*	*Variable 2 (RA-C)*
Mean	470.3666667	283.8333333
Variance	4958.654023	742.6954023
Observations	30	30
Pearson Correlation	0.144931896	
Hypothesised Mean Difference	0	
df	29	
t Stat	14.24363378	
P(T<=t) one-tail	6.33026E-15	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	1.26605E-14	
t Critical two-tail	2.045229642	

Group A FC	*Variable 1 (STR E-C)*	*Variable 2 (RA-E)*
Mean	491.8666667	304.2333333
Variance	7541.016092	1043.84023
Observations	30	30
Pearson Correlation	0.29446761	
Hypothesised Mean Difference	0	
df	29	
t Stat	12.34312645	
P(T<=t) one-tail	2.27691E-13	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	4.55381E-13	
t Critical two-tail	2.045229642	

Group C FC	*Variable 1 (STR E-G)*	*Variable 2(RA-E)*
Mean	504.6	256.35
Variance	6668.673684	513.5026316
Observations	20	20
Pearson Correlation	0.444566297	
Hypothesised Mean Difference	0	
df	19	
t Stat	14.9201676	
P(T<=t) one-tail	3.02274E-12	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	6.04548E-12	
t Critical two-tail	2.093024054	

Table 7.2b*: Paired Two-Sample T-Test for Means [Temporal EVS of Group B at S-I and S-T in RA-E]

T-Test: Paired Two Sample for Means

Group B Temporal EVS in RA-E	*at S-I*	*at S-T*
Mean	1036.557143	837.190127
Variance	12916.25193	32738.76854
Observations	35	35
Pearson Correlation	0.410944097	
Hypothesised Mean Difference	0	
df	34	
t Stat	6.955654032	
P(T<=t) one-tail	2.54392E-08	
t Critical one-tail	1.690924255	
P(T<=t) two-tail	5.08785E-08	
t Critical two-tail	2.032244509	

Table 7.2c*: Paired Two-sample T-tests for Means [Temporal EVS of Group C at S-I and S-T in RA-E and STR E-G]

T-Test: Paired Two Sample for Means

Group C Temporal EVS	*RA-E*		*STR E-G*	
	at S-I	*at S-T*	*at S-I*	*at S-T*
Mean	1085.363333	846.355	2471.14	1961.689306
Variance	14452.69777	17461.67945	434098.4909	148808.4625
Observations	20	20	20	20
Pearson Correlation	0.770637071		0.77713269	
Hypothesised Mean Difference	0		0	
df	19		19	
t Stat	12.40073151		5.256341893	
P(T<=t) one-tail	7.42355E-11		2.24949E-05	
t Critical one-tail	1.729132812		1.729132812	
P(T<=t) two-tail	1.48471E-10		4.49898E-05	
t Critical two-tail	2.093024054		2.093024054	

Table 7.2d*: Paired Two-sample T-tests for Means [Temporal EVS of Group A at S-I and S-T in RA-C, RA-E, STR C-E, and STR E-C]

T-Test: Paired Two Sample for Means

Group A Temporal EVS	*RA-E*		*STR E-C*	
	at S-I	*at S-T*	*at S-I*	*at S-T*
Mean	1181.64	861.62	2166.495185	1559.316481
Variance	14244.09007	6113.14197	389020.9299	102809.8621
Observations	30	30	30	30
Pearson Correlation	0.294560903		0.578720014	
Hypothesised Mean Difference	0		0	
df	29		29	
t Stat	14.37905381		6.517676044	
P(T<=t) one-tail	4.97227E-15		1.94508E-07	
t Critical one-tail	1.699127027		1.699127027	
P(T<=t) two-tail	9.94454E-15		3.89016E-07	
t Critical two-tail	2.045229642		2.045229642	
Group A Temporal EVS	*RA-C*		*STR C-E*	
	at S-I	*at S-T*	*at S-I*	*at S-T*
Mean	1187.136111	724.7027778	1909.118148	1213.685202
Variance	20195.74113	10666.09037	271981.8971	74895.01209
Observations	30	30	30	30
Pearson Correlation	0.179137378		0.090774353	
Hypothesized Mean Difference	0		0	
df	29		29	
t Stat	15.82924386		6.723364467	
P(T<=t) one-tail	4.15827E-16		1.11843E-07	
t Critical one-tail	1.699127027		1.699127027	
P(T<=t) two-tail	8.31654E-16		2.23686E-07	
t Critical two-tail	2.045229642		2.045229642	

Table 7.3a*: Two-sample T-tests Assuming Equal Variances [Temporal EVS of Group A and C at S-I and S-T in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Group A S-I	*Group C S-I*	*Group A S-T*	*Group C S-T*
Mean	1181.64	1085.363333	861.62	846.355
Variance	14244.09007	14452.69777	6113.14197	17461.67945
Observations	30	20	30	20
Pearson Correlation	14326.66395		10605.27139	
Hypothesised Mean Difference	0		0	
df	48		48	
t Stat	2.78637222		0.513483554	
P(T<=t) one-tail	0.003805539		0.304984353	
t Critical one-tail	1.677224196		1.677224196	
P(T<=t) two-tail	0.007611078		0.609968706	
t Critical two-tail	2.010634758		2.010634758	

Table 7.3b*: Two-sample T-tests Assuming Equal/ Unequal Variances [Temporal EVS of Group A and B at S-I and S-T in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	Group A S-I	*Group B S-I*
Mean	1181.64	1036.557143
Variance	14244.09007	12916.25193
Observations	30	35
Pearson Correlation	13527.47901	
Hypothesised Mean Difference	0	
df	63	
t Stat	5.013553301	
P(T<=t) one-tail	2.30927E-06	
t Critical one-tail	1.669402222	
P(T<=t) two-tail	4.61854E-06	
t Critical two-tail	1.998340543	

T-Test: Two-Sample Assuming Unequal Variances

	Group A S-T	*Group B S-T*
Mean	861.62	837.190127
Variance	6113.14197	32738.76854
Observations	30	35
Hypothesised Mean Difference	0	
df	48	
t Stat	0.723815591	
P(T<=t) one-tail	0.236345987	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	0.472691975	
t Critical two-tail	2.010634758	

Table 7.3c*: Two-sample T-tests Assuming Equal Variances [Temporal EVS of Group B and C at S-I and S-T in RA-E]

T-Test: Two-Sample Assuming Equal Variances

	at S-I		*at S-T*	
	Group C	*Group B*	*Group C*	*Group B*
Mean	1085.363333	1036.557143	846.355	837.190127
Variance	14452.69777	12916.25193	17461.67945	32738.76854
Observations	20	35	20	35
Pearson Correlation	13467.05327		27262.07622	
Hypothesised Mean Difference	0		0	
df	53		53	
t Stat	1.500397379		0.198022557	
P(T<=t) one-tail	0.069722796		0.421892499	
t Critical one-tail	1.674116237		1.674116237	
P(T<=t) two-tail	0.139445592		0.843784999	
t Critical two-tail	2.005745995		2.005745995	

Table 7.3d*: Paired Two-sample T-tests for Means [Temporal EVS of Group C at S-I and S-T in RA-E and STR E-G]

T-Test: Paired Two Sample for Means

	at S-I		*at S-T*	
	STR E-G	*RA-E*	*STR E-G*	*RA-E*
Mean	2471.14	1085.363333	1961.689306	846.355
Variance	434098.4909	14452.69777	148808.4625	17461.67945
Observations	20	20	20	20
Pearson Correlation	0.308272963		0.103747496	
Hypothesised Mean Difference	0		0	
df	19		19	
t Stat	9.802402853		12.64111501	
P(T<=t) one-tail	3.62578E-09		5.35189E-11	
t Critical one-tail	1.729132812		1.729132812	
P(T<=t) two-tail	7.25157E-09		1.07038E-10	
t Critical two-tail	2.093024054		2.093024054	

Table 7.3e*: Paired Two-sample T-tests for Means [Temporal EVS of Group A at S-I and S-T in RA-E and STR E-C]

T-Test: Paired Two Sample for Means

	at S-I		*at S-T*	
	STR E-C	*RA-E*	*STR E-C*	*RA-E*
Mean	2166.495185	1181.64	1559.316481	861.62
Variance	389020.9299	14244.09007	102809.8621	6113.14197
Observations	30	30	30	30
Pearson Correlation	-0.021909525		-0.261388463	
Hypothesised Mean Difference	0		0	
df	29		29	
t Stat	8.460350659		10.93945908	
P(T<=t) one-tail	1.26714E-09		4.14199E-12	
t Critical one-tail	1.699127027		1.699127027	
P(T<=t) two-tail	2.53427E-09		8.28397E-12	
t Critical two-tail	2.045229642		2.045229642	

Table 7.3f*: Paired Two-sample T-tests for Means [Temporal EVS of Group A at S-I and S-T in RA-C and STR C-E]

T-Test: Paired Two Sample for Means

	at S-I		*at S-T*	
	STR C-E	*RA-C*	*STR C-E*	*RA-C*
Mean	1909.118148	1187.136111	1213.685202	724.7027778
Variance	271981.8971	20195.74113	74895.01209	10666.09037
Observations	30	30	30	30
Pearson Correlation	0.321966798		0.136089531	
Hypothesised Mean Difference	0		0	
df	29		29	
t Stat	7.998143127		9.597839673	
P(T<=t) one-tail	4.02516E-09		8.30437E-11	
t Critical one-tail	1.699127027		1.699127027	
P(T<=t) two-tail	8.05031E-09		1.66087E-10	
t Critical two-tail	2.045229642		2.045229642	

Table 7.4a*: Paired Two-sample T-tests for Means [Temporal EVS of Group A at S-I and S-T in STR E-C and STR C-E]

T-Test: Paired Two Sample for Means

EVS at S-I	*STR E-C*	*STR C-E*
Mean	2166.495185	1909.118148
Variance	389020.9299	271981.8971
Observations	30	30
Pearson Correlation	0.455744629	
Hypothesised Mean Difference	0	
df	29	
t Stat	2.334926414	
P(T<=t) one-tail	0.013338291	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	0.026676582	
t Critical two-tail	2.045229642	
EVS at S-T	*STR E-C*	*STR C-E*
Mean	1559.316481	1213.685202
Variance	102809.8621	74895.01209
Observations	30	30
Pearson Correlation	0.399627271	
Hypothesised Mean Difference	0	
df	29	
t Stat	5.772001404	
P(T<=t) one-tail	1.4912E-06	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	2.98239E-06	
t Critical two-tail	2.045229642	

Table 7.4b*: Two-sample T-tests Assuming Equal Variances [Temporal EVS Rates of Increase of Group A and C at S-I and S-T from RA-E to STR E-C and STR E-G]

T-Test: Two-Sample Assuming Equal Variances

Temporal EVS Rate of Increase at S-I	*Group C*	*Group A*
Mean	1.283764535	0.8504369
Variance	0.321245866	0.31199034
Observations	20	30
Pearson Correlation	0.315653988	
Hypothesised Mean Difference	0	
df	48	
t Stat	2.67178416	
P(T<=t) one-tail	0.005136748	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	0.010273496	
t Critical two-tail	2.010634758	
Temporal EVS Rate of Increase at S-T	*Group C*	*Group A*
Mean	1.363258109	0.83020427
Variance	0.30094952	0.18328083
Observations	20	30
Pearson Correlation	0.229858017	
Hypothesised Mean Difference	0	
df	48	
t Stat	3.851517487	
P(T<=t) one-tail	0.000173737	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	0.000347473	
t Critical two-tail	2.010634758	

Table 7.4c*: Two-sample T-tests Assuming Equal Variances [TFD, FC and FSD Rates of Increase of Group A and C from RA-E to STR E-C and STR E-G]

T-Test: Two-Sample Assuming Equal Variances

TFD Rate of Increase	*From RA-E to STR E-G*	*From RA-E to STR E-C*
Mean	1.179412517	0.563459131
Variance	0.141845127	0.066063777
Observations	20	30
Pooled Variance	0.096060561	
Hypothesised Mean Difference	0	
df	48	
t Stat	6.884397052	
P(T<=t) one-tail	5.5297E-09	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	1.10594E-08	
t Critical two-tail	2.010634758	
FC Rate of Increase	*From RA-E to STR E-G*	*From RA-E to STR E-C*
Mean	0.971587517	0.625507833
Variance	0.084959268	0.085312904
Observations	20	30
Pooled Variance	0.085172923	
Hypothesised Mean Difference	0	
df	48	
t Stat	4.107863003	
P(T<=t) one-tail	7.73491E-05	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	0.000154698	
t Critical two-tail	2.010634758	
FSD Rate of Increase	*From RA-E to STR E-G*	*From RA-E to STR E-C*
Mean	1.108781651	0.601426207
Variance	0.095498438	0.062990089
Observations	20	30
Pooled Variance	0.075857977	
Hypothesised Mean Difference	0	
df	48	
t Stat	6.381199447	
P(T<=t) one-tail	3.27134E-08	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	6.54268E-08	
t Critical two-tail	2.010634758	

Bibliography

Abrams, S. G., and Zuber, B. L. 1972. "Some Temporal Characteristics of Information Processing during Reading." *Reading Research Quarterly* 8 (1): 42-51.

Adab, Beverly. 2000. "Evaluating Translation Competence." In *Developing Translation Competence*, edited by Christina Schäffner and Beverly Adab, 215-28. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Agrifoglio, Marjorie. 2003. "El uso de la traducción a vista en la enseñanza de la Interpretación." *Tradução & Comunicação* 12 (1): 99-113.

Agrifoglio, Marjorie. 2004. "Sight Translation and Interpreting: A Comparative Analysis of Constraints and Failures." *Interpreting* 6 (1): 43-67.

Ahrens, August. 1891. *Untersuchungen über die Bewegung der Augen beim Schreiben*. Rostok: Boldt.

Allport, D. Alan, Barbara Antonis, and Patricia Reynolds. 1972. "On the Division of Attention: A Disproof of the Single Channel Hypothesis." *Quarterly Journal of Experimental Psychology* 24 (2): 225-235.

Al-Qurashi, Abdulwali. 2004. *Problems of Sight Translation*. MA Thesis. Jordan: Yarmouk University.

Alves, Fábio, José Luiz Gonçalves, and Karina Szpak. 2012. "Identifying Instances of Processing Effort in Translation through Heat Maps: An Eye-tracking Study Using Multiple Input Sources." Paper presented at the 24th International Conference on Computational Linguistics, Mumbai, 15th Dec 2012.

Alves, Fábio, Adriana Pagano, and Igor da Silva. 2009. "A New Window on Translators' Cognitive Activity: Methodological Issues in the Combined Use of Eye Tracking, Key Logging and Retrospective Protocols." In *Methodology Technology and Innovation in Translation Process Research: A Tribute to Arnt Lykke Jakobsen*, edited by Inger M. Mees, Fabio Alves, and Susanne Göpferich, 268-269. Copenhagen: Samfundslitteratur.

Anderson, Irving Howard, and Walter F. Dearborn. 1952. *The Psychology of Teaching Reading*. New York: The Ronald Press.

Anderson, John. 2000. *Cognitive Psychology and Its Implications.* 5th edition. New York: Worth.

Angelelli, Claudia. 1999. "The Role of Reading in Sight Translation." *The ATA Chronicle (Translation Journal of the American Association of Translators)* 28 (5): 27-30.

Ashby, Jane, Keith Rayner, and Charles Clifton. 2005. "Eye Movements of Highly Skilled and Average Readers: Differential Effects of Frequency and Predictability." *The Quarterly Journal of Experimental Psychology* 58 (6): 1065-1086.

Baddeley, Alan. 1986. *Working Memory*. Oxford: Clarendon Press.

Balling, Laura W., Kristian T. Hvelplund, and Annette C. Sjørup. 2014. "Evidence of Parallel Processing during Translation." *Meta: Translators' Journal* 59 (2): 234-259.

Bangalore, Srinivas, Bergljot Behrens, Michael Carl, Maheshwar Ghankot, Arndt Heilmann, Jean Nitzke, Moritz Schaeffer, and Annegret Sturm. 2016. "Syntactic Variance and Priming Effects in Translation." In *New Directions in Empirical Translation Process Research*, edited by Michael Carl, Srinivas Bangalore, and Moritz Schaeffer, 211-238. New York: Springer.

Bayer-Hohenwarter, Gerrit. 2010. "Comparing Translational Creativity Scores of Students and Professionals: Flexible Problem-Solving and/or Fluent Routine Behaviour?" In *New Approaches in Translation Process Research*, edited by Susanne Göpferich, Fabio Alves, and Inger M. Mees, 83-112. Frederiksberg: Samfundslitteratur.

Beatty, Jackson. 1982. "Task-evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources." *Psychological Bulletin* 91 (2): 276-292.

Bond, Guy Loraine, and Miles Albert Tinker. 1973. *Reading Difficulties: Their Diagnosis and Correction*. New York: Appleton-Century-Crofts.

Book, William Frederick. 1908. *The Psychology of Skill: With Special Reference to Its Acquisition in Typewriting*. Missoula: University of Montana.

Bouma, H., and A. H. De Voogd. 1974. "On the Control of Eye Saccades in Reading." *Vision Research* 14 (4): 273-284.

Braze, David, Tao Gong, and Hosung Nam. 2018. "Reading Fluency Matters: NIH R21 HD090460-01A1." *PsyArXiv*. 2nd January 2018. https://psyarxiv.com/b563t/.

Breitmeyer, Bruno G. 1983. "Sensory Masking, Persistence, and Enhancement in Visual Exploration and Reading." In *Eye Movements in Reading*, edited by Keith Rayner, 3-29. New York and London: Academic Press.

Broadbent, Donald E. 1958. *Perception and Communication*. London: Pergammon Press.

Buswell, Guy Thomas. 1920. *An Experimental Study of the Eye-Voice Span in Reading*. Chicago: University of Chicago Press.

Buswell, Guy Thomas. 1921. “The Relationship between Eye-perception and Voice-Response in Reading.” *Journal of Educational Psychology* 12 (4): 217-227.

Buswell, Guy Thomas. 1927. *A Laboratory Study of the Reading of Modern Foreign Languages*. New York: The Macmillan Co.

Buswell, Guy Thomas. 1937. *How Adults Read*. Chicago: University of Chicago Press.

Butterworth, Brian. 1980. “Evidence from Pauses in Speech.” In *Language Production*, edited by Brian Butterworth, 155-176. New York and London: Academic Press.

Caffrey, Colm. 2008. “Using Pupillometric, Fixation-based and Subjective Measures to Measure the Processing Effort Experienced when Viewing Subtitled TV Anime with Pop-up Gloss.” In *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*, edited by Arnt Lykke Jakobsen, Susanne Göpferich, and Inger M. Mees, 125-144. Copenhagen: Samfundslitteratur.

Campbell, Fergus W., and Robert H. Wurtz. 1979. “Saccadic Omission: Why We Do not See a Gray-out during a Saccadic Eye Movement.” *Vision Research* 18 (10): 1297-1301.

Campbell, Stuart. 1998. *Translation into the Second Language*. London and New York: Longman.

Carl, Michael, and Barbara Dragsted. 2012. “Inside the Monitor Model: Processes of Default and Challenged Translation Production.” *Translation: Corpora, Computation, Cognition. Special Issue on the Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation* 2 (1): 127-145.

Carl, Michael, and Moritz Schaeffer. 2016. “The Eye-Key Span as a Measure for Translation Complexity.” The Scandinavian Workshop on Applied Eye Tracking, SWAET 2016, University of Turku, Finland. 19th-21st June 2016.

Carpenter, Patricia A., and Marcel Adam Just. 1983. “What Your Eyes Do while Your Mind is Reading.” In *Eye Movements in Reading*, edited by Keith Rayner, 275-307. New York and London: Academic Press.

Carver, Ronald P. 1982. “Optimal Rate of Reading Prose.” *Reading Research Quarterly* 18 (1): 56-88.

Carver, Ronald P. 1992. “Reading Rate: Theory, Research, and Practical Implications.” *Journal of Reading* 36 (2): 84-95.

Castelhano, Monica S., and John M. Henderson. 2008. "Stable Individual Differences across Images in Human Saccadic Eye Movements." *Canadian Journal of Experimental Psychology* 62 (1): 1-14.

Chafe, Wallace, and Jane Danielewicz. 1987. "Properties of Spoken and Written Language." In *Comprehending Oral and Written Language*, edited by Rosalind Horowitz and S. Jay Samuels, 83-113. New York and London: Academic Press.

Chang, Vincent Chieh-ying. 2009. *Testing Applicability of Eye-tracking and fMRI to Translation and Interpreting Studies: An Investigation into Directionality.* PhD thesis. London: Imperial College.

Chmiel, Agnieszka, and Iwona Mazur. 2013. "Eye Tracking Sight Translation Performed by Trainee Interpreters." In *Tracks and Treks in Translation Studies*, edited by Catherine Way, Sonia Vandepitte, Reine Meylaerts, Magdalena Bartłomiejczyk, 189-205. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Chmiel, Agnieszka, Agnieszka Szarkowska, Danijel Koržinek, Agnieszka Lijewska, Łukasz Dutka, Łukasz Brocki, and Krzysztof Marasek. 2017. "Ear-Voice Span and Pauses in Intra- and Interlingual Respeaking: An Exploratory Study into Temporal Aspects of the Respeaking Process." *Applied Psycholinguistics* 38 (5): 1201-1227.

Christoffels, Ingrid, and Annette M. B. De Groot. 2009. "Simultaneous Interpreting." In *Handbook of Bilingualism: Psycholinguistic Approaches*, edited by Judith F. Kroll and Annette M. B. De Groot, 454-479. Oxford: Oxford University Press.

Clement, Jesper, and Henrik Selsøe Sørensen. 2008. "Do Average Consumers Read and Understand Food Labels? Outline of a Pilot Study." In *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*, edited by Arnt Lykke Jakobsen, Susanne Göpferich, and Inger M. Mees, 145-155. Copenhagen: Samfundslitteratur.

Clifton, Charles. 1983. "Psycholinguistic Factors Reflected in the Eye." In *Eye Movements in Reading*, edited by Keith Rayner, 269-272. New York and London: Academic Press.

Craik, John William. 1948. "Theory of the Human Operator in Control Systems." *British Journal of Psychology* 38 (3): 142-148.

Crawley, Michael J. 2002. *Statistical Computing: An Introduction to Data Analysis Using S-Plus*. Chichester: John Wiley and Sons.

De Luca, Maria, Maria Pontillo, Silvia Primativo, Donatella Spinelli, and Pierluigi Zoccolotti. 2013. "The Eye-voice Lead during Oral Reading in Developmental Dyslexia."

Frontiers in Human Neuroscience. 6th November 2013. https://doi.org/10.3389/fnhum.2013.00696.

Deignan, Alice. 2005. *Metaphor and Corpus Linguistics*. *Vol. 6: Converging Evidence in Language and Communication Research.* Amsterdam and Philadelphia: John Benjamins Publishing Company.

Dejean le Féal, Karla. 1981. "L'enseignement des méthodes d'interprétation." *L'enseignement de l'interprétation et de la traduction*, *Cahiers de traductologie* 4 (1): 380-403.

Deutsch, Alin, and Diana Deutsch. 1963. "Attention: Some Theoretical Considerations." *Psychological Review* 70 (1): 8-90.

Dodge, Raymond, and Thomas Sparks Cline. 1901. "The Angle Velocity of Eye Movements." *Phychological Review* 8 (2): 145-157.

Downey, June E., and John E. Anderson. 1915. "Automatic Writing." *American Journal of Psychology* 26 (1): 161-195.

Downing, John A., and Che Kan Leong. 1982. *Psychology of Reading*. New York: Macmillan.

Dragsted, Barbara. 2004. *Segmentation in Translation and Translation Memory Systems: An Empirical Investigation of Cognitive Segmentation and Effects of Integrating a TM System into the Translation Process.* Ph.D. Series 5. Copenhagen: Copenhagen Business School.

Dragsted, Barbara. 2005. "Segmentation in Translation: Differences across Level of Expertise and Difficulty." *Target* 17 (1): 49-70.

Dragsted, Barbara. 2010. "Coordination of Reading and Writing Processes in Translation." In *Translation and Cognition*, edited by Gregory M. Shreve and Erik Angelone, 41-62. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Dragsted, Barbara, and Inge Gorm Hansen. 2007. "Speaking your Translation: Exploiting Synergies between Translation and Interpreting." In *Interpreting Studies and Beyond: A Tribute to Miriam Shlesinger*, edited by Fran Pöchhacker, Arnt Lykke Jakobsen, and Inger M. Mees, 251–274. Copenhagen: Samfundslitteratur.

Dragsted, Barbara, and Inge Gorm Hansen. 2008. "Comprehension and Production in Translation: A Pilot Study on Segmentation and the Coordination of Reading and Writing Processes." In *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*, edited by Arnt Lykke Jakobsen, Susanne Göpferich, and Inger M. Mees, 9-29. Copenhagen: Samfundslitteratur.

Dragsted, Barbara, and Inge Gorm Hansen. 2009. “Exploring Translation and Interpreting Hybrids. The Case of Sight Translation.” *Meta: Translators' Journal* 54 (3): 588-604.

Dragsted, Barbara, Inge Gorm Hansen, and Henrik Selsøe Sørensen. 2009. “Experts Exposed.” In *Methodology Technology and Innovation in Translation Process Research: A Tribute to Arnt Lykke Jakobsen*, edited by Inger M. Mees, Fabio Alves, and Susanne Göpferich, 294-317. Copenhagen: Samfundslitteratur.

Dragsted, Barbara, Inger Margrethe Mees, and Inge Gorm Hansen. 2011. “Speaking Your Translation: Students' First Encounter with Speech Recognition Technology.” *Translation & Interpreting* 3 (1): 10-43.

Duchowski, Andrew. 2003. *Eye-tracking Methodology: Theory and Practice*. New York: Springer.

Ehrensberger-Dow, Maureen, and Gary Massey. 2013. “Indicators of translation competence: Translators' Self-concepts and the Translation of Titles.” *Journal of Writing Research* 5 (1): 103-131.

Ehrlich, Kate. 1983a. “Eye Movements in Pronoun Assignment: A Study of Sentence Integration.” In *Eye Movements in Reading,* edited by Keith Rayner, 253-268. New York and London: Academic Press.

Ehrlich, Susan F. 1983b. “Contextual Influences on Eye Movement in Reading.” In *Eye Movements in Reading,* edited by Keith Rayner, 193-201. New York and London: Academic Press.

Ehrlich, Susan F., and Keith Rayner. 1981. “Contextual Effects on Word Perception and Eye Movements during Reading.” *Journal of Verbal Learning and Verbal Behavior* 20 (6): 641-655.

Ericsson, K. Anders. 2010. “Expertise in interpreting.” In *Tranlation and Cognition*, edited by Gregory Shreve and Erik Angelone, 231-262. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Ericsson, K. Anders, and Herbert A. Simon. 1980. “Verbal Reports as Data.” *Psychological Review* 87 (3): 215-251.

Evans, Vyvyan, Benjamin K. Bergen and Jörg Zinken. 2007. “The Cognitive Linguistics Enterprise: An Overview.” In *The Cognitive Linguistics Readers*, edited by Vyvyan Evans, Benjamin K. Bergen, and Jörg Zinken, 2-36. London and Oakville: Equinox Publishing.

Fairbanks, Grant. 1937. “The Relation between Eye-movements and Voice in Oral Reading of Good and Poor Readers.” *Psychological Monographs* 48 (3): 78-107.

Ferreira F., and John M. Henderson. 1990. “Use of Verb Information during Syntactic Parsing: Evidence form Eye-tracking and Word by Word Self-paced Reading.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16 (4): 555-568.

Fillmore, Charles J. 1975. “An Alternative to Checklist Theories of Meaning.” *Processing of the First Annual Meeting of the Berkeley Linguistics Society* 1: 123-131.

Fine, Philip, Anna Berry, and Burton Rosner. 2006. “The Effect of Pattern Recognition and Tonal Predictability on Sight-Singing Ability.” *Psychology of Music* 34 (4): 431-447.

Frazier, Lyn. 1983. “Processing Sentence Structure.” In *Eye Movements in Reading*, edited by Keith Rayner, 215-236. New York and London: Academic Press.

Gaarder, Kenneth R. 1970. “Eye Movements and Perception.” In *Early Experience and Visual Information Processing in Perceptual and Reading Disorders*, edited by Francis A. Young and B. Lindsley Donald, 79-94. Washington, D.C.: National Academy of Sciences.

Gerloff, Pamela Ann. 1988. *From French to English: A Look at the Translation Process in Students, Bilinguals, and Professional Translators*. PhD Thesis. Cambridge, Massachusetts: Harvard University.

Geyer, John Jacob. 1966. *Perceptual Systems in Reading: A Temporal Eye-Voice Span Constant*. PhD Thesis. Berkeley: University of California.

Gibson Eleanor J., and Harry Levin. 1975. *The Psychology of Reading*. Cambridge, Massachusetts: The MIT Press.

Gile, Daniel. 1984. “Les noms propres en interprétation simultanée.” *Multilingua-Journal of Cross-Cultural and Interlanguage Communication* 3 (2): 79-86.

Gile, Daniel. 1997. “Conference Interpreting as a Cognitive Management Problem.” In *Cognitive Processes in Translation and Interpreting*, edited by Joseph H. Danks, Gregory M. Shreve, Stephen B. Fountain, and Michael K. McBeath, 196-214. London: Sage.

Gile, Daniel. 2009. *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Godijn, Richard, and Jan Theeuwes. 2003. “The Relationship between Exogenous and Endogenous Saccades and Attention.” In *The Mind’s Eye: Cognitive and Applied*

Aspects of Eye Movement Research, edited by Jukka Hyönä, Ralph Radach, and Heiner Deubel, 3-26. Amsterdam: Elsevier Science.

Goolsby. Thomas. W. 1994. "Eye Movement in Music Reading: Effects of Reading Ability, Notational Complexity, and Encounters." *Music Perception* 12 (1): 77-96.

Göpferich, Susanne. 2010. "The Translation of Instructive Texts from a Cognitive Perspective: Novices and Professionals Compared." In *New Approaches in Translation Process Research*, edited by Susanne Göpferich, Fabio Alves, and Inger M. Mees, 5-56. Frederiksberg: Samfundslitteratur.

Göpferich, Susanne, and Riitta Jääskeläinen. 2009. "Process Research into the Development of Translation Competence: Where are We, and Where do We Need to Go?" *Across Languages and Cultures* 10 (2): 169-191.

Göpferich, Susanne, Arnt Lykke Jakobsen, and Inger M. Mees, eds. 2008. *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur.

Gorszczyńska, Paula. 2010. "The Potential of Sight Translation to Optimize Written Translation: The Example of the English-Polish Language Pair." *Selected Papers of the CETRA Research Seminar in Translation Studies 2009*, 1-12.

Gray, Clarence Truman.1917. *Types of Reading Ability as Exhibited through Tests and Laboratory Experiments: An Investigation Subsidized by the General Education Board*. Chicago: University of Chicago Press.

Grosman, Meta. 2000. "Non-mother Tongue Translation: An Open Challenge." In *Translation into Non-mother Tongues: In Professional Practice and Training*, edited by Meta Grosman, Mira Kadric, Irena Kovacic, and Mary Snell-Hornby, 21-34. Tübingen: Stauffenburg Verlag.

Hess, Eckhard H., and James M. Polt. 1964. "Pupil Size in Relation to Mental Activity during Simple Problem-solving." *Science* 143 (3611): 1190-1192.

Hirst, William, Elizabeth S. Spelke, Celia C. Reaves, George Caharack, and Ulric Neisser. 1980. "Dividing Attention without Alternation of Automaticity." *Journal of Experimental Psychology* 109 (1): 98-117.

Hoffman, Robert R. 1997. "The Cognitive Psychology of Expertise and the Domain of Interpreting." *Interpreting* 2 (1): 189-230.

Hogaboam, Thomas W. 1983. "Reading Patterns in Eye Movement Data." In *Eye Movements* *in Reading*, edited by Keith Rayner, 309-332. New York and London: Academic Press.

Holmqvist, Kenneth, Marcus Nystrom, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures.* Oxford: Oxford University Press.

Huang, Chih-Chieh. 2011. *Tracking Eye Movements in Sight Translation- The Comprehending Process in Interpreting*. Master Thesis. Taibei: National Taiwan Normal University.

Hubel, David H., and Torsten N. Wiesel. 1962. "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex." *Journal of Physiology* 160 (1): 106-145.

Huey, Edmund Burke. 1898. "Preliminary Experiments in the Physiology and Psychology of Reading." *American Journal of Psychology* 9 (4): 575-586.

Huey, Edmund Burke. 1908. *The Psychology and Pedagogy of Reading*. Oxford: Macmillan.

Hvelplund, Kristian Tangsgaard. 2011. *Allocation of Cognitive Resources in Translation: An Eye-tracking and Key-logging Study.* PhD thesis. Copenhagen: Copenhagen Business School.

Hvelplund, Kristian Tangsgaard. 2014. "Eye Tracking and the Translation Process: Reflections on the Analysis and Interpretation of Eye Tracking Data." *MonTI Special Issue-Minding Translation* 2014: 201-223.

Hvelplund, Kristian Tangsgaard. 2015. "Four Fundamental Types of Reading during Translation." In *Translation in Transition: Between Cognition, Computing and Technology,* edited by Arnt Lykke Jakobsen and Bartolomé Mesa-Lao, 55-80. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Hyönä, Jukka, Robert F. Lorch, and Mike Rinck. 2003. "Eye Movement Measures to Study Global Text Processing." *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* 2003: 313-334.

Hyönä, Jukka, Ralph Radach, and Heiner Deubel. 2003. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam: Elsevier.

Hyönä, Jukka, Jorma Tommola, and Anna-Mari Alaja. 1995. "Pupil Dilation as a Measure of Processing Load in Simultaneous Interpretation and Other Language Tasks." *Quarterly Journal of Experimental Psychology* 48 (3): 598-612.

Ilg, Gérard, and Sylvie Lambert. 1996. "Teaching Consecutive Interpreting." *Interpreting* 1

(1): 69-99.

Inhoff, Albrecht Werner. 1983. “Attentional Strategies during the Reading of Short Stories.” In *Eye Movements in Reading*, edited by Keith Rayner, 181-192. New York and London: Academic Press.

Inhoff, Albrecht Werner, and Keith Rayner. 1986. “Parafoveal Word Processing during Eye Fixations in Reading: Effects of Word Frequency.” *Perception & Psychophysics* 40 (6): 431–439.

Inhoff, Albrecht. W., Matthew Solomon, Ralph Radach, and Bradley A. Seymour. 2011. “Temporal Dynamics of the Eye Voice Span and Eye Movement Control during Oral Reading.” *Journal of Cognitive Psychology* 23 (5): 543-558.

Iqbal, Shamsi T., Piotr D. Adamcyzk, Xianjun S. Zheng, and Brian P. Bailey. 2005. “Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution.” Paper presented at the Conference on Human Factors in Computing Systems, Portland, Oregon, USA, 2nd-7th April 2005.

Jääskeläinen, Riitta. 1999. *Tapping the Process. An Explorative Study of the Cognitive and Affective Factors Involved in Translating*. Publications in the Humanities 22. Joensuu: University of Joensuu.

Jacobson, O. Irving. 1941. “An Analytical Study of Eye Movements in Reading Vocal and Instrumental Music.” *Journal of Musicology* 3 (1): 3-32.

Jaffe, Joseph, Stanley Feldstein, and L. A. Cassota. 1967. “A Stochastic Model of Speaker Switching in Natural Dialogue.” In *Research in Verbal Behavior and some Neurological Implications*, edited by K. Salzinger and S. Salzinger, 281-294. New York and London: Academic Press.

Jakobsen, Arnt Lykke. 2003. “Effects of Think Aloud on Translation Speed, Revision and Segmentation.” In *Triangulating Translation*, edited by Fabio Alves, 69-95. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Jakobsen, Arnt Lykke. 2011. “Tracking Translators’ Keystrokes and Eye Movements with Translog.” In *Methods and Strategies of Process Research. Integrative Approaches in Translation Studies*, edited by C. Alvstad, A. Hild and E. Tiselius, 37-55. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Jakobsen, Arnt Lykke, and Kristian T. H. Jensen. 2008. “Eye Movement Behaviour across Four Different Types of Reading Task.” In *Looking at Eyes: Eye-tracking Studies of*

Reading and Translation Processing, edited by Arnt Lykke Jakobsen, Susanne Göpferich, and Inger M. Mees, 101-124. Copenhagen: Samfundslitteratur.

Järvilehto, Timo, Veli-Matti Nurkkala, Kyösti Koskela, Elina Holappa, and Heidi Vierelä. 2008. "Reading as Anticipatory Formation of Meaning: Eye Movement Characteristics and Fixation-speech Intervals when Articulating Different Types of Text." *Journal of Transfigural Mathematics* 1 (1): 73-81.

Jegerski, Jill. 2013. "Self-paced Reading." In *Research Methods in Second Language Psycholinguistics*, edited by Jill Jegerski and Bill VanPatten, 20-49. New York and London: Routledge.

Jensen, Christian. 2008. "Assessing Eye-tracking Accuracy in Translation Studies." In *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*, edited by Arnt Lykke Jakobsen, Susanne Göpferich, and Inger M. Mees, 157-174. Copenhagen: Samfundslitteratur.

Jensen, Kristian Tangsgaard Hvelplund, Annette C. Sjørup, and Laura W. Balling. 2009. "Effects of L1 Syntax on L2 Translation." In *Methodology Technology and Innovation in Translation Process Research: A Tribute to Arnt Lykke Jakobsen*, edited by Inger M. Mees, Fabio Alves, Susanne Göpferich, 319-336. Copenhagen: Samfundslitteratur.

Judd, Charles Hubbard, and Guy Thomas Buswell. 1922. *Silent Reading: A Study of the Various Types*. Chicago: University of Chicago.

Jungermann, Helmut, Hans-Rüdiger Pfister, and Katrin Fischer. 2010. *Die Psychologie der Entscheidung*. Spektrum, Akad: Verlag.

Just, Marcel A., and Patricia A. Carpenter. 1980. "A Theory of Reading: From Eye Fixations to Comprehension." *Psychological Review* 87 (4): 329-354.

Keating, Gregory D. 2013. "Eye-tracking With Text." In *Research Methods in Second Language Psycholinguistics*, edited by Jill Jegerski and Bill VanPatten, 69-92. New York and London: Routledge.

Kim, Changmin. 2001. *Sight Translation in Its Own Right*. Master Thesis. Monterey Bay: California State University.

Kolers, Paul A. 1976. "Buswell's Discoveries." In *Eye Movements and Psychological Processes*, edited by Richard A. Monty and John W. Senders, 371-395. Hillsdale, New Jersey: Erlbaum.

Korpal, Paweł. 2012. "On language-pair Specificity in Sight Translation: An Eye-Tracking

Study." Paper Presented at the 2nd Internationale Fachkonferenz des BDÜ-Übersetzens in die Zukunft, Berlin, Germany. 28th-30th September 2012.

Krieber, Magdalena, Katrin D. Bartl-Pokorny, Florian B. Pokorny, Dajie Zhang, Karin Landerl, Christof Körner, Franz Pernkopf, Thomas Pock, Christa Einspieler, and Peter B. Marschik. 2017. "Eye Movements during Silent and Oral Reading in a Regular Orthography: Basic Characteristics and Correlations with Childhood Cognitive Abilities and Adolescent Reading Skills." *PLoS ONE*. 2nd February 2017. https://doi.org/10.1371/journal.pone.0170986.

Krings, Hans P. 1988."Blick in die 'Black Box'–Eine Fallstudie zum Übersetzungsprozeß bei Berufsübersetzern." In *Textlinguistik und Fachsprache. Akten des Internationalen übersetzungswissenschaftlichen AILA-Symposions, Hildesheim*, edited by R. Arntz, 393–412. Hildesheim: Olms.

Kroll, Judith F., and Erika Stewart. 1994. "Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations." *Journal of Memory and Language* 33 (2): 149-174.

Lakoff, George. 1990. "The Invariance Hypothesis: Is Abstract Reason Based on Image-schemas?" *Language* 54 (3): 610-646.

Lakoff, George, and H. Thompson. 1975. "Introduction to Cognitive Grammar." *Proceedings of the 1st Annual Meeting of the Berkeley Linguistics Society*, 295-313. Berkeley, CA.: Berkeley Linguistics Society.

Lambert, Sylvie. 2004. "Shared Attention during Sight Translation, Sight Interpretation and Simultaneous Interpretation." *Meta: Translators' Journal* 49 (2): 294-306.

Läubli, Samuel, and Ulrich Germann. 2016. "Statistical modelling and automatic tagging of human translation processes." In *New Directions in Empirical Translation Process Research,* edited by Michael Carl, Srinivas Bangalore, and Moritz Schaeffer, 155-181. New York: Springer.

Laubrock, Jochen, and Christiane Bohn 2008. "Broadening the Scope of Eye-movement Research in Reading: Oral Reading and Proof Reading." *International Journal of Psychology* 43 (3): 190.

Laubrock, Jochen, and Reinhold Kliegl. 2015. "The Eye-voice Span during Reading *Aloud." Frontiers in Psychologe*. 24th September 2015. https://doi.org/10.3389/fpsyg.2015.01432.

Laubrock, Jochen, Reinhold Kliegl, and R. Engbert. 2007. “On the Eye-voice Span during Oral Reading.” Paper Presented at the European Conference on Eye Movements, Potsdam, Germany. 9th-23rd August 2007.

Lawson, Everdina A. 1961. “A Note on the Influence of Different Orders of Approximation to the English Language upon Eye-voice Span.” *Quarterly Journal of Experimental Psychology* 13 (1): 53-55.

Ledbetter, Frances Gresham. 1947. “Reading Reactions for Varied Types of Subject Matter: An Analytic Study of the Eye Movements of Eleventh Grade Pupils.” *The Journal of Educational Research* 41 (2): 102-115.

Lee, Jieun. 2012. “What Skills Do Student Interpreters Need to Learn in Sight Translation Training?” *Meta: Translators’ Journal* 57 (3): 694-714.

Levin, Harry, and Ann Buckler Addis. 1979. *The Eye-voice Span*. Cambridge, Massachusetts: The MIT Press.

Levin, Harry, and Julie A. Cohn. 1968. *Studies of Oral Reading: The Effects of Instructions on the Eye-voice Span*. Ithaca, New York: Cornell University.

Levin, Harry, and Eleanor. L. Kaplan. 1968. “Eye-Voice Span (EVS) within Active and Passive Sentences.” *Language and Speech* 11 (4): 251-258.

Levin, Harry, and Ann Turner. 1968. “Sentence Structure and the Eye-Voice Span.” In *The Analysis of Reading Skill: A Program of Basic and Applied Research*, edited by Harry Levin, Eleanor J. Gibson, and James J. Gibson, 196-220. New York: Office of Education, Bureau of Research.

Li, Xiangdong. 2014. “Sight Translation as a Topic in Interpreting Research: Progress, Problems, and Prospects.” *Across Languages and Cultures* 15 (1): 67-89.

Liu, Yanmei, Binghan Zheng, and Hao Zhou. 2018. “Measuring the Difficulty of Text Translation: The Combination of Text-focused and Translator-oriented Approaches.” *Target*. 8th January 2019. https://doi.org/10.1075/target.18036.zhe.

Liversedge, Simon P., and John M. Findlay. 2000. “Saccadic Eye Movements and Cognition.” *Trends in Cognitive Science* 4 (1): 6-14.

Macizo, Pedro, and M. Teresa Bajo. 2004. “When Translation Makes the Difference: Sentence Processing in Reading and Translation.” *Psicológica* 25 (1): 181-205.

Marmaridou, Sophia. 1996. “Directionality in Translation Processes and Practices.” *Target* 8 (1): 49-73.

Marshall, Sandra P., C. W. Pleydell-Pearce, and B. T. Dickson. 2003. "Integrating Psychological Measures of Cognitive Workload and Eye Movements to Detect Strategy Shifts." Paper presented at the 36th Hawaii International Conference on System Sciences, HI, USA. 6th-9th January 2003.

Martin, Ann. 1993. "Teaching Sight Translation to Future Interpreters." In *Translation: The Vital Link,* edited by Catriona Picken, 398-405. London: Institute of Translation and Interpreting.

McConkie. George W. 1981. "Evaluating and Reporting Data Quality in Eye Movement Research." *Behaviour Research Methods* 13 (2): 97-106.

McConkie, George W. 1983. "Eye Movements and Perception during Reading." In *Eye Movements in Reading*, edited by Keith Rayner, 65-95. New York and London: Academic Press.

McConkie, George W., and Keith Rayner. 1975. "The Span of the Effective Stimulus during a Fixation in Reading." *Perception and Psychophysics* 17 (6): 578-586.

McConkie, George W., and Shun Nan Yang. 2003. "How Cognition Affects Eye Movements during Reading." In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, edited by Jukka Hyönä, Ralph Radach, and Heiner Deubel, 413-427. Amsterdam: Elsevier Science.

Morris, Robin, and Rihana Williams. 2004. "Eye Movements, Word Familiarity, and Vocabulary Acquisition." In *Eye movements and Information Processing during Reading. A Special Issue of the European Journal of Cognitive Psychology*, edited by Alan Kennedy, Ralph Radach, and Keith Rayner, 312-339. Hove: Psychology Press.

Morrison, Robert E. 1983. "Retinal Image Size and the Perceptual Span in Reading." In *Eye Movements in Reading*, edited by Keith Rayner, 31-40. New York and London: Academic Press.

Morton, John. 1964. "The Effects of Context upon Speed of Reading, Eye Movements and Eye-voice Span." *Quarterly Journal of Experimental Psychology* 16 (4): 340-354.

Moser-Mercer, Barbara. 1995. "Sight Translation and Human Information Processing." In *Basic Issues in Translation Studies,* edited by Albrecht Neubert, Gregory Shreve, and Klaus Gommlich, Kent, 159-166. O.H.: Institute for Applied Linguistics.

Moser-Mercer, Barbara, U. H. Frauenfelder, Beatriz Casado, and Alexander Kunzli. 2000. "Searching to Define Expertise in Interpreting." In *Language Processing and*

Simultaneous Interpreting. Interdisciplinary Perspective, edited by B. Englund-Dimitrova and K. Hyltenstam, 1-21. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Nakayama, Minoru, Koji Takahashi, and Yasutaka Shimizu. 2002. "The Act of Task Difficulty and Eye-movement Frequency for the 'Oculo-motor Indices'." In *Proceedings of the 2002 Symposium on Eye-Tracking Research & Applications*, 37-42. New York: ACM.

Neisser, Ulric. 1967. *Cognitive Psychology.* New York: Appleton-Century-Crofts.

Neubert, Albrecht. 2000. "Competence in Language, in Languages, and in Translation." In *Developing Translation Competence*, edited by Christina Schäffner and Beverly Adab, 4-18. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Newmark, Peter. 1988. *A Textbook of Translation*. Hemel Hempstead: Prentice Hall International.

Nilsen, Anne Birgitta, and May-Britt Monsrud. 2015. "Reading Skills for Sight Translation in Public Sector Services." *Translation & Interpreting* 7 (3): 10-20.

Norman, Donald A. 1968. "Toward a Theory of Memory and Attention." *Psychological Review* 75 (6): 522-36.

O'Brien, John Anthony. 1926. *Reading, Its Psychology and Pedagogy*. New York: The Century Company.

O'Brien, Sharon. 2006. "Eye-tracking and Translation Memory Matches." *Perspectives* 14 (3): 185-205.

O'Brien, Sharon. 2008. "Processing Fuzzy Matches in Translation Memory Tools: An Eye-Tracking Analysis." In *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*, edited by Arnt Lykke Jakobsen, Susanne Göpferich, and Inger M. Mees, 79-102. Copenhagen: Samfundslitteratur.

O'Brien, Sharon. 2009. "Eye Tracking in Translation Process Research: Methodological Challenges and Solutions." In *Methodology Technology and Innovation in Translation Process Research: A Tribute to Arnt Lykke Jakobsen*, edited by Inger M. Mees, Fabio Alves, Susanne Göpferich, 251-266. Copenhagen: Samfundslitteratur.

Olsen, Anneli. 2012. *The Tobii I-VT Fixation Filter.* Tobii Technology. https://stemedhub.org/resources/2173/download/Tobii_WhitePaper_TobiiIVTFixationFilter.pdf.

O'Regan, Kevin. 1979. "Saccade Size Control in Reading: Evidence for the Linguistic Control Hypothesis." *Perception and Psychophysics* 25 (6): 501-509.

Orozco, Mariana. 2000. "Building A Measuring Instrument for the Acquisition of Translation Competence in Trainee Translators." In *Developing Translation Competence*, edited by Christina Schäffner and Beverly Adab, 199-214. Amsterdam and Philadelphia: John Benjamins Publishing Company.

PACTE. [Beeby, Allison, Mauro Fernández Rodríguez, Olivia Fox, Amparo Hurtado Albir, Wilhelm Neunzig, Mariana Orozco, Marisa Presas, Patricia Rodríguez Inés, and Lupe Romero. (Principal Investigator: Amparo Hurtado Albir)]. 2003. "Building a Translation Competence Model." In *Triangulating Translation: Perspectives in process Oriented Research*, edited by Fábio Alves, 43-68. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Pavlović, Nataša. 2007. "Directionality in Translation and Interpreting Practice." *FORUM. International Journal of Interpretation and Translation* 5 (2): 79-99.

Pavlović, Nataša, and Kristian T. H. Jensen. 2009. "Eye Tracking Translation Directionality." In *Translation Research Projects 2*, edited by Anthony Pym and Alexander Perekrestenko, 93-109. Tarragona: Intercultural Studies Group.

Pavlović, Tanja. 2014. "Exploring Directionality in Translation Studies." *Explorations in English Language and Linguistics* 1 (2): 149-165.

Pedersen, Viggo Hjørnager. 2000. "Translation into L2 – In practice, and in the classroom." In *Translation into Non-mother Tongues*, edited by Meta Grosman, Mira Kadric, Irena Kovacic, and Mary Snell-Hornby, 109-116. Tübingen: Stauffenburg Verlag.

Pedersen, Emilie Sand, and Helle V. Dam. 2014. "Short-Term Memory in the Production Phase of Sight Translation." *Hermes-Journal of Language and Communication in Business* 27 (52): 93-105.

Pellicer-Sánchez, Ana. 2016. "Incidental L2 Vocabulary Acquisition from and while Reading: An Eye-tracking Study." Studies in *Second Language Acquisition* 38 (1): 97-130.

Penfield, Wilder, and Lamar Roberts. 1959. *Speech and Brain Mechanisms*. Princeton, N.J.: Princeton University Press.

Pernice, Kara, and Jakob Nielsen. 2009. *Eye-tracking Methodology- How to Conduct and Evaluate Usability Studies Using Eye-tracking*. Berkeley, C.A.: New Riders Press.

Pickering, Martin J., Steven Frisson, Brian McElree, and Matthew Traxler. 2004. "Eye

Movements and Semantic Composition." In *The On-line Study of Sentence Comprehension: Eye-tracking, ERPs, and Beyond*, edited by Manuel Carreiras and and Carreiras Clifton Jr., 33-50. Hove: Psychology Press.

Pöchhacker, Franz. 2004. *Introducing Interpreting Studies*. New York and London: Routledge.

Pokorn, Nike. 2005. *Challenging the Traditional Axioms: Translation into a Non-mother Tongue*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Posner, Michael I. 1980. "Orienting of attention." *Quarterly Journal of Experimental Psychology* 32 (1): 3-25.

Prassl, Friedrike. 2010. "Translator's Decision-making Processes in Research and Knowledge Integration." In *New Approaches in Translation Process Research*, edited by Susanne Göpferich, Fabio Alves, and Inger M. Mees, 57-82. Frederiksberg: Samfundslitteratur.

Presas, Marisa. 2000. "Bilingual Competence and Translation Competence." In *Developing Translation Competence*, edited by Christina Schäffner and Beverly Adab, 19-31. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Quantz, John Oscar. 1897. *Problems in the Psychology of Reading*. New York: Macmillan.

Rayner, Keith. 1978. "Eye Movements in Reading and Information Processing." *Psychological Bulletin* 85 (3): 618-660.

Rayner, Keith. 1983. *Eye Movements in Reading*. New York and London: Academic Press.

Rayner, Keith. 1986. "Eye Movements and the Perceptual Span in Beginning and Skilled Readers." *Journal of Experimental Child Psychology* 41 (2): 211–236.

Rayner, Keith. 1995. "Eye Movements and Cognitive Processes in Reading, Visual Search, and Scene Perception." In *Eye Movement Research: Mechanisms*, Processes and Applications, edited by John Malcolm Findlay, Robin Walker, and Robert W. Kentridge, 3-22. Netherlands: Elsevier Science.

Rayner, Keith. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124 (3): 372-422.

Rayner, Keith. 2009. "Eye Movements and Attention in Reading, Scene Perception, and Visual Search." *The Quarterly Journal of Experimental Psychology* 62(8): 1457-1506.

Rayner, Keith, and George W. McConkie. 1976. "What Guides a Reader's Eye Movements?"

Vision Research 16 (8): 829-837.

Rayner, Keith, and Alexander Pollatsek. 1989. *The Psychology of Reading*. Englewood Cliffs. N. J.: Prentice Hall.

Rayner, Keith, Arnold D. Well, and Alexander Pollatsek. 1980. "Asymmetry of the Effective Visual Field in Reading." *Perception and Psychophysics* 27 (6): 537–544.

Rayner, Keith, Kathryn H. Chace, Timothy J. Slattery, and Jane Ashby. 2006. "Eye Movements as Reflections of Comprehension Process in Reading." *Scientific Studies of Reading* 10 (3): 241-255.

Rayner, Keith, Xingshan Li, Carrick C. Williams, Kyle R. Cave, and Arnold D. Well. 2007. "Eye Movements during Information Processing Tasks: Individual Differences and Cultural Effects." *Vision Research* 47 (21): 2714-2726.

Resnick, Lauren B. 1970. "Relations between Perceptual and Syntactic Control in Oral Reading." *Journal of Educational Psychology* 61 (5): 382-385.

Rosch, Eleanor. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104 (3):192-233.

Russo, J. Edward, Eric J. Johnson, and Debra L. Stephens. 1989. "The Validity of Verbal Protocols." *Memory & Cognition* 17 (6): 759-769.

Saldanha, Gabriela, and Sharon O'Brien. 2014. *Research Methods in Translation Studies*. London and New York: Routledge.

Sampaio, Glória Regina Loreto. 2007. "Mastering Sight Translation Skills." *Tradução & Comunicação* 16 (1): 63-9.

Sandrelli, Annalisa. 2003."New Technologies in Interpreter Training: CAIT." In *Textologie und Translation*, edited by Heidrun Gerzymisch-Arbogast, 261-294. Tübingen: Gunter Narr Verlag.

Schaeffer, Moritz, Barbara Dragsted, Kristian T. Hvelplund, Laura W. Balling, and Michael Carl. 2016. "Word Translation Entropy: Evidence of Early Target Language Activation during Reading for Translation." In *New Directions in Empirical Translation Process Research*, edited by Michael Carl, Srinivas Bangalore, and Moritz Schaeffer, 155-181. New York: Springer.

Schäffner, Christina, and Beverly Adab, eds. 2000. *Developing Translation Competence*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Schilperoord, Joost. 1996. *It's about Time: Temporal Aspects of Cognitive Processes in Text Production*. Utrecht: Rodopi.

Schlesinger, Izchak M. 1968. *Sentence Structure and the Reading Process*. The Hague: Mouton.

Schmaltz, Márcia, Igor AL da Silva, Adriana Pagano, Fabio Alves, Ana Luísa V. Leal, Derek F. Wong, Lidia S. Chao, and Paulo Quaresma. 2016. "Cohesive Relations in Text Comprehension and Production: An Exploratory Study Comparing Translation and Post-editing." In *New Directions in Empirical Translation Process Research*, edited by Michael Carl, Srinivas Bangalore, and Moritz Schaeffer, 239-263. New York: Springer.

Schnipke, Susan K., and Marc W. Todd. 2000. "Trials and Tribulations of Using an Eye-tracking System." In *CHI'00 Extended Abstracts on Human Factors in Computing Systems*, 273-274. New York: ACM Press.

Seeber, Killian G. 2007. "Processing German Verb-final Structures in Simultaneous Interpreting: TEPR Evidence." Paper presented at the International Conference on Processing Head-final Structures, Rochester Institute of Technology, Rochester, New York. 21st-22nd September 2007.

Seleskovitch, Danica. 1976. "Interpretation: A Psychological Approach to Translating." In *Translation: Applications and Research*, edited by R. Anderson, W. Bruce, and Richard W. Brislin, 92-116. New York: Gardner.

Setton, Robin, and Manuela Motta. 2007. "Syntacrobatics: Quality and Reformulation in Simultaneous-with-text." *Interpreting* 9 (2): 199-230.

Shaffer, L. Henry, and J. Hardwick. 1969. "Reading and typing." *The Quarterly Journal of Experimental Psychology* 21 (4): 381-383.

Sharmin, Selina, Oleg Spakov, Kari-Jouko Räihä, and Arnt Lykke Jakobsen. 2008 "Where on the Screen do Translation Students Look While Translating, and for How Long?" In *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*, editd by Arnt Lykke Jakobsen, Susanne Göpferich, and Inger M. Mees, 31-51. Copenhagen: Samfundslitteratur.

Shebilske, Wayne. 1975. "Reading Eye Movements from an Information-processing Point of View." In *Understanding Language*, edited by D. W. Massaro, 291-311. New York and London: Academic Press.

Shebilske, Wayne L., and Dennis F. Fisher. 1983. "Eye Movements and Context Effects during Reading of Extended Discourse." In *Eye Movements in Reading*, edited by Keith Rayner, 153-179. New York and London: Academic Press.

Shlesinger, Miriam, and Noam Ordan. 2012. "More Spoken or More Translated? Exploring a Known Unknown of Simultaneous Interpreting." *Target* 24 (1): 43-60.

Shreve, Gregory M., and Bruce J. Diamond. 1997. "Cognitive Processes in Translation and Interpreting: Critical Issues." In *Cognitive Processes in Translation and Interpreting*, edited by Joseph H. Danks, G. H. Shreve, S. B. Fountain, and M. K. Mcbeath, 233-251. Thousand Oaks, C.A.: Sage Publications.

Shreve, Gregory M., Isabel Lacruz, and Erik Angelone. 2010. "Cognitive Effort Syntactic Disruption and Visual Interference in a Sight Translation Task." In *Translation and Cognition*, edited by Gregory M. Shreve and Erik Angelone, 63-84. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Silva, Susana, and São Luís Castro. 2018. "The Time Will Come: Evidence for an Eye-Audiation Span in Silent Music Reading." *Sage Journals*. 10th April 2018. https://doi.org/10.1177/0305735618765302.

Sjørup, Annette Camilla. 2008. "Metaphor Comprehension in Translation: Methodological Issues in a Pilot Study." In *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*, edited by Arnt Lykke Jakobsen, Susanne Göpferich, and Inger M. Mees, 53-77. Copenhagen: Samfundslitteratur.

Sjørup, Annette Camilla. 2013. *Cognitive Effort in Metaphor Translation: An Eye-tracking and Key-logging Study*. PhD thesis. Copenhagen: Copenhagen Business School.

Sloboda. John A. 1985. *The Musical Mind*. Oxford: Clarendon Press.

Slowiaczek, Maria L. 1983. "What Does the Mind Do while the Eyes Are Gazing." In *Eye Movements in Reading*, edited by Keith Rayner, 345-356. New York and London: Academic Press.

Smith, Frank. 1971. *Understanding Reading*. New York: Holt, Rinehart, and Winston.

Solomons, Leon M., and Gertrude Stein. 1896. "Normal Motor Automatism." *Psychological Review* 3 (5): 492-512.

Song, Zhongwei. 2010. "Skill Transfer from Sight Translation to Simultaneous Interpreting: A Case Study of Effective Teaching Technique." *International Journal of Interpreter Education* 2 (2010): 120-134.

Spelke, Elizabeth S., William C. Hirst, and Ulric Neisser. 1976. "Skills of Divided Attention." *Cognition* 4 (3): 215-230.

Sperling, George. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs: General and Applied* 74 (11): 1-29.

Stansfield, Charles W. 2008. *A Practical Guide to Sight Translation of Assessments*. Rockville, MD: Second Language Testing, Inc.

Starr, Matthew S., and Keith Rayner. 2001. "Eye Movements during Reading: Some Current Controversies." *Trends in Cognitive Sciences* 5 (4): 156-63.

Stuart-Hamilton, Ian, and Pat Rabbitt. 1997. "The Decline of Eye-voice Span in Elderly Readers." *Educational Gerontology* 23 (4): 389-400.

Sunderman, Gretchen. 2013. "Translation Recognition Tasks." In *Research Methods in Second Language Psycholinguistics*, edited by Jill Jegerski and Bill VanPatten, 185-211. New York and London: Routledge.

Tabachnick, Barbara G., and Linda S. Fidell. 2000. *Using Multivariate Statistics*. Boston: Allyn and Bacon.

Thompson, W. Burt. 1987. "Music Sight-reading Skill in Flute Players." *The Journal of General Psychology* 114 (4): 345-352.

Timarová, Šárka. 2010. "Interpreting Process Research: Paradigms and Methods." In *New Approaches in Translation Process Research*, edited by Susanne Göpferich, Fabio Alves, and Inger M. Mees, 133-150. Frederiksberg: Samfundslitteratur.

Timarová, Šárka, Ivana Čeňková, Reine Meylaerts, Erik Hertog, Arnaud Szmalec, and Wouter Duyck. 2015. "Simultaneous Interpreting and Working Memory Capacity." In *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting*, edited by Aline Ferreira and John W. Schwieter, 101-126. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Tinker, Miles A. 1951. "Fixation Pause Duration in Reading." *The Journal of Educational Research* 44 (6): 471-479.

Tinker, Miles A. 1965. *Bases for Effective Reading*. Minneapolis: University of Minnesota Press.

Van Orden, Karl F., Tzyy-Ping Jung, and Scott Makeig. 2000. "Combined Eye Activity Measures Accurately Estimate Changes in Sustained Visual Task Performance." *Biological Psychology* 52 (3): 221-240.

Vargas-Urpi, Mireia. 2018. "Sight translation in Public service interpreting: a dyadic or triadic exchange?" *The Interpreter and Translator Trainer* (2018): 1-17.

Vernon, Magdalen Dorothea. 2014. *The Experimental Study of Reading*. Cambridge: Cambridge University Press.

Vienne, Jean. 2000. "Which Competence Should We Teach to Future Translators, and How." In *Developing Translation Competence*, edited by Christina Schäffner and Beverly Adab, 91-100. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Viezzi, Mario. 1990. "Sight Translation, Simultaneous Interpretation and Information Retention." In *Aspects of Applied and Experimental Research on Conference Interpretation*, edited by Laura Gran and Christopher Taylor, 54-60. Udine: Campanotto.

Volkmann, Frances C. 1962. "Vision during Voluntary Saccadic Eye Movements." *Journal of the Optical Society of America* 52 (5): 571-578.

Walker, Callum. 2018. "A Cognitive Perspective on Equivalent Effect: Using Eye Tracking to Measure Equivalence in Source Text and Target Text Cognitive Effects on Reader's." *Perspectives*. 22nd March 2018. https://doi.org/10.1080/0907676X.2018.1449870.

Wanat, Stanley F. 1971. "Linguistic Structure in Reading: Models from the Research of Project Literacy." In *The Literature of Research in Reading with Emphasis on Models*, edited by Frederick B. Davis, 604-626. New Brunswick, N.J.: Rutgers, the State University.

Wanat, Stanley F. 1976. "Relations between Language and Visual Processing." In *Theoretical Models and Processes of Reading*, edited by Harry Singer and Robert B. Ruddell, 108-137. Newark, Del.: IRA.

Wang, Yifang. 2017. *Translating Linguistic Metaphors in Both Directions: A Process-Oriented Study on English-Chinese Translation*. PhD Thesis. Durham: Durham University.

Weber, Wilhelm K. 1990. "The Importance of Sight Translation in an Interpreter Training Program." *ATA Scholarly Monograph Series* 4 (1): 44-52.

Welford, Alan T. 1952. "The 'Psychological Refractory Period' and the Timing of High-speed Performance-a Review and a Theory." *British Journal of Psychology* 43 (1): 2-19.

Wolverton, Gary S. and David Zola. 1983. "The Temporal Characteristics of Visual

Information Extraction during Reading." In *Eye Movements in Reading*, edited by Keith Rayner, 41-51. New York and London: Academic Press.

Zheng, Binghan. 2012a. *Choice-making in the Process of English-Chinese Translation: An Empirical Study*. Beijing: Foreign Language Teaching and Research Press.

Zheng, Binghan. 2012b. "The Role of Consultation Sources Revisited: An Empirical Study of English–Chinese Translation." *Perspectives* 22 (1): 113-135.

Zheng, Binghan, and Xia Xiang. 2013. "Processing Metaphorical Expressions in Sight Translation: An Empirical-experimental Research." *Babel* 59 (2): 160-183.

Zheng, Binghan, and Xia Xiang. 2014. "The Impact of Cultural Background Knowledge in the Processing of Metaphorical Expressions: An Empirical Study of English-Chinese Sight Translation." *Translation and Interpreting Studies* 9 (1): 5-24.

Zheng, Binghan, and Hao Zhou. 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.

Zollinger, Ruth Harold. 1974. *The Psychological Reality of Information Focus for the Reader*. PhD Thesis. Cleveland, O.H.: Case Western Reserve University.