Ampersand 11 (2023) 100131

Contents lists available at ScienceDirect

Ampersand

journal homepage: www.elsevier.com/locate/amper

Cognitive processing of the extra visual layer of live captioning in simultaneous interpreting. Triangulation of eye-tracked process and performance data

Lu Yuan *, Binhua Wang

Centre for Translation Studies, University of Leeds, UK

ARTICLE INFO

Keywords:
Live captioning
Cognitive processing
Eye-tracking
Triangulation

ABSTRACT

While real-time automatic captioning has become available on various online meeting platforms, it poses additional cognitive challenges for interpreters because it adds an extra layer for information processing in interpreting. Against this background, this empirical study investigates the cognitive processing of live captioning in interpreting on Zoom Meetings. 13 interpreting trainees in a postgraduate professional training programme were recruited for an eye-tracking experiment of simultaneous interpreting under two conditions: with live captioning on and with live captioning off. Their eye movement data and interpreting performance data were collected during the experiment. Three questions were explored: 1) How do the interpreters process the additional layer of visual information from live captioning? 2) Which types of information segments tax more cognitive resources in interpreting with live captioning? 3) Is there a significant difference in interpreting accuracy between interpreting with live captioning and interpreting without live captioning? The results showed the following findings: 1) Although participants were observed to constantly shift their attention between the live transcript area and the non-live transcript area, they tended to consciously keep their visual attention to the live captioning area when numbers and proper names appeared. 2) With live captioning on, it required more cognitive effort to process the information containing a higher density of numbers and proper names than processing information without numbers and proper names. 3) There was a significant improvement in the number and proper name accuracy in interpreting with live captioning.

1. Introduction

Multimodality is a distinctive feature of simultaneous interpreting (SI). Traditionally, interpreters have to process verbal (what the speaker said), paraverbal information (the pauses, stress, intonation, speed, prosody, articulation, fluency, or hesitation in the speech) and nonverbal information (visual perception) (Wang, 2018). However, since the pandemic, there has been a growing number of online meeting platforms. The live captioning option offered on these platforms has posed additional cognitive challenges for interpreters. Gile (2009, p. 182) indicated in his "tightrope hypothesis" that simultaneous interpreters normally work at the limit of their processing capacity. The multimodal processing of live captioning alongside with verbal, paraverbal and nonverbal information has challenged the cognitive processing capacity of simultaneous interpreters to a new level. Prior cognitive interpreting studies have largely focused on the processing of the three channels of information. For instance, the source speech difficulty (Kuang and Zheng, 2022) and listening comprehension (Hyönä et al., 1995) for the verbal channel; prosodic features such as delivery rate (Korpal and Stachowiak-Szymczak, 2020) and intonation (Shlesinger, 1994) for the paraverbal channel; visual perception of gestures (Vranjes and Brône, 2021) and face (Seeber, 2017) for the nonverbal channel. To this date, however, no empirical research has touched upon the cognitive processing of live captioning. Given the lack of studies on live captioning, the primary aim of this study is to make an initial attempt to fill in the gap by exploring live captioning as an additional layer of information processed in interpreting.

Although fruitful results have been yielded in the processing pattern of subtitles from the audio-visual research (Kruger and Steyn, 2014). Some efforts have also been made to investigate the processing of transcription in the simultaneous interpreting (Seeber et al., 2020). Live captioning, however, has nuanced differences from the subtitles and

* Corresponding author. Centre for Translation Studies, University of Leeds, LS2 9JT, UK.
E-mail address: mllyu@leeds.ac.uk (L. Yuan).

https://doi.org/10.1016/j.amper.2023.100131
Received 22 February 2023; Received in revised form 17 May 2023; Accepted 20 June 2023
Available online 20 June 2023

transcription. Subtitles and transcription are fully congruent with what the speaker said, while live captioning is not error-free. Some of the information such as numbers and proper names can be quite accurate with live captioning, but there still exist incongruences between live captioning and audio input due to the limitation of current voice recognition technology. Previous literature reported on how the (in) congruence between what is presented and what has been said affects the cognitive processing pattern of simultaneous interpreters (Stachowiak-Szymczak and Korpal, 2019; Chmiel et al., 2020). Nonetheless, the studies mentioned employed static experimental stimuli, such as slides and manuscripts, which inherently differ from the dynamic nature of live captioning. When utilizing live captioning, interpreters are faced with the dual challenge of processing the dynamic nature of the text and engaging in critical evaluation of its accuracy.

While it is indeed plausible that the utilization of text from live captioning can potentially enhance the overall completeness and accuracy of interpretation, it is important to acknowledge the existence of a trade-off between the benefits derived from using the text and the potential distraction caused by live captioning. Furthermore, the additional cognitive effort required to process the extra layer of visual information further complicates the matter.

Against this background, the second aim of this study is to investigate the potential processing pattern of live captions in interpreting and how it might affect interpreting performance.

2. Literature review

2.1. Studying multimodality in simultaneous interpreting from a cognitive perspective

The term multimodality refers to "the use of several semiotic modes in the design of a semiotic product or event" (Kress and Van Leeuwen, 2001, p. 20). The concept of multimodality was coined to denote a non-hierarchal relationship between several semiotic modes: verbal, vocal and kinesic modalities (Müller et al., 2013). Poyatos (1983) also pointed out that verbal language, paralanguage and kinesics are indispensable triple structures for communication. This means these three information channels are interrelated and are of equal importance for simultaneous interpreters to process. Simultaneous interpreters are multitaskers who not only process information from verbal and para-verbal channels but also have visual access to speakers' faces or gestures in the booth. Bühler (1985) similarly considered interpreting as a "multichannel communication phenomenon", they launched a survey investigation to explore the role of visual information in nonverbal communication to AIIC (International Association of Conference Interpreters) professional interpreters. Their results reported that most of the interpreters considered the visual information of the speakers significant for their comprehension. In recent years, research on nonverbal communication has been a strong focus on multimodality in simultaneous interpreting. Galhano Rodrigues and Isabel (2007) made an attempt to explore the gestures of simultaneous interpreters in a real-life conference setting based on her qualitative microanalysis model of speech and body movement, the findings proved that gestures played a cognitive role during SI. Galvão (2013) adopted a mixed design to explore speech gestures in SI among professionals, their results indicated the gestures served pragmatic, discursive and cognitive functions.

In addition to the visual perception of gestures, simultaneous interpreters often have access to visual aids such as PowerPoints in conferences. Given the fact that nonverbal information is radically different from verbal information, it might be more cognitively demanding for interpreters to process. Hence, cognition has become a major trend in multimodalities in simultaneous interpreting (Pöchhacker, 2021).

To explore multimodality in simultaneous interpreting from a cognitive perspective, a widely known conceptual framework the Effort Model by Daniel Gile (2009) allowed us to explore multimodal processing from a cognitive perspective. He drew from cognitive psychology's concept of "energy" and formed his concept of "efforts" for the interpreting process. Based on this model, multiple channels of information compete for attentional resources which makes simultaneous interpreters normally work at the limit of their cognitive processing capacity. Gile listed several variants from a traditional SI model: SI = L (Listening comprehension) + M (Memory) + P (Production)+ C (Coordination) (Gile, 2009). Additional input of visual information requires extra Reading (R) effort which cooperates with the Listening effort but also competes for interpreters' limited processing capacity (Gile, 2009, p. 182). These multimodal inputs that interpreters have to process require a different amount of cognitive effort from interpreters, which might affect the interpreting quality to a large extent. Previous experimental studies on multimodality in SI have largely focused on the effect of visual input on interpreting performance. Chmiel et al. (2020) made an attempt to test whether the (in)congruence between verbal and nonverbal channel affect the interpreting performance. They manipulated 60 items comprising proper names, numbers and content words as the experimental stimuli. In the congruent condition, visual information comprising proper names, numbers and content words fully matched the verbal speech. In the incongruent condition, they prepared different surnames, altered the digits, and replaced content words with other similar words of the same length. Although their results did not prove a facilitatory effect of visual input on the overall interpreting quality, the accuracy of numbers was higher in the congruent condition. Indicated by longer viewing times and lower accuracy, they also suggested that additional incongruent visual input triggers higher cognitive load as interpreters were forced to invest more reading effort to resolve the conflict between visual and verbal channels. Stachowiak-Szymczak and Korpal (2019) designed an experiment in which both interpreting trainees and professionals were tested in two conditions: with and without slides. Their results reported an increasing accuracy of numbers with slides. These findings suggested the visual modality does not conflict with verbal modality but instead played a facilitative role as indicated by higher interpreting accuracy. The facilitative role of visual information on interpreting quality has also been proposed in other attempts (Korpal and Stachowiak, 2013; Lambert, 2004; Stachowiak, 2017).

The studies listed above have mainly focused on the effect of visual information on interpreting performance. However, it is not enough to reveal a complete picture of the multimodal nature of simultaneous interpreting. The investigation of how interpreters process multiple inputs is of equal importance. In fact, most of the attention in process-oriented research in interpreting studies has been awarded to the simultaneous interpreting (Ahrens, 2017). This can be explained from the view that simultaneous interpreters normally work under a lot of stress and the overlapping of different tasks in SI makes it distinctive for investigating cognitive processing.

Seeber (2012) made it one step further by examining the multimodal input in SI through various modalities, for instance, verbal-visual and visual-spatial modalities. Instead of focusing on the effect of different modalities on interpreting performance, he focused more on how interpreters processed these multimodal inputs (such as slides, faces and gestures). With his innovative design, he also proposed that with the appropriate experiment design, eye-tracking holds the potential to explore the interpreting process.

2.2. Exploring the cognitive processing pattern in simultaneous interpreting with eye tracking

Just and Carpenter (1980) suggested that "there is no appreciable lag between what is being fixated and what is being processed" from their Eye-Mind Hypothesis. Based on this hypothesis, where the interpreters fixate on indicates what they are processing. The eye-tracking technology allows real-time measurements (Holmqvist et al., 2011) during SI along with limited invasiveness (Seeber, 2020). Given its nature, eye-tracking is a feasible method to explore the cognitive processing

pattern in simultaneous interpreting.

Previous experimental studies have set out to explore the cognitive processing pattern in SI from the following three aspects. One area of this research aims to explore preferential attention patterns when processing different channels during SI tasks. Few attempts have been made based on the concept of the ear-voice span (EVS), which has been considered a fundamental measurement for cognitive processing in simultaneous interpreting (Timarova et al., 2011). Few attempts have been made to explore the preferential attention patterns when processing different channels during SI tasks. Seeber et al. (2020) prepared a video recording with transcripts as experimental stimuli. Interpreters were assigned two tasks: one is simultaneous interpreting, and the control task was reading while listening. Instead of drawing each word as an area of interest (AOI), they chose to identify AOIs at the sentence level: the previous sentence, the critical phrase of the sentence, the rest of the sentence and the following sentence. Mean dwell time and fixation proportions were analysed, their findings showed interpreters different processing patterns in two tasks. Interpreters tend to follow the visual channel more than the verbal channel in the RWL task as indicated by higher fixation proportions observed in the rest of the sentence and they tended to look ahead of the verbal channel. However, when performing simultaneous interpreting, interpreters tend to follow verbal channels more than visual channels as a clear visual lag has been observed. Their findings pointed to an ear-lead eye processing pattern, which did not lend support to the assumption that simultaneous interpreters normally follow a predictive processing pattern. This contradicts the results from Amos et al. (2022) who adopted a visual search paradigm in their eye-tracking experiment and proved the prediction behaviour during simultaneous interpreting. Recently, Seeber (2020)'s findings have been supported by Zou et al. (2022) Interpreters were asked to perform simultaneous interpreting with the source text presented on the screen. To investigate attention patterns, they innovatively created the concept of ear-eye span (EIS) which refers to "the temporal difference between the source speech input and the source written text input." They calculated the ear-voice span (EVS), eye-voice span (IVS) and ear-eye span (EIS) along with the assessment of the interpreting performance, and they observed that the average EIS is −4 seconds**.** This indicates that in general interpreters start to process the written text in the visual channel 4 s later than listening to the audio speech.

The second line of research focuses on the processing pattern in the visual channel. Nonverbal information encompassed various forms of visual inputs, to name a few, the face, gestures and manuscripts. Do interpreters follow a certain processing pattern when faced with various forms of visual information? Previous literature has drawn different conclusions. Seeber (2012) set up an experiment when interpreters are exposed to multiple visual information: speakers' face, gestures and slides containing numbers. Each type of visual information was created as an AOIs. He compared the gaze duration among these AOIs, the results reported a gaze duration on speakers' faces than on slides while the gestures attracted the least attention among the three. In addition, the gaze pattern observed on the slide point to the fact that interpreters were searching for visual information about numbers on the slides when they hear the number from the verbal channel. Seeber (2012) explained this can be seen as a clue to prove that simultaneous interpreters were actively searching for visual information which might be complementary to verbal information. In a more recent study, Serbert (2019) incorporated the most comprehensive visual information. Unlike the previous studies which were arranged in a laboratory, this experiment was performed in a more authentic environment in the booth. Participants, therefore, were exposed to not only the slides and the speaker but also the conference room and the audience. Among various types of visual information, the results revealed that interpreters rely on visual information on slides the most. A clear processing pattern was observed, interpreters seek visual information on the slide once it is presented. It corroborated with Seeber (2012) that interpreters were actively searching for visual information that might facilitate the comprehension of verbal information.

The third line of these research areas stems from the problem triggers**.** According to the Effort Model from Gile (2009), simultaneous interpreters normally work at the limit of their processing capacity. Problems triggers such as numbers and proper names are "associated with increased processing capacity requirements or cause attention management problems" (Gile 2019, p. 171). To testify this hypothesis, cognitive processing studies in simultaneous interpreting have mainly engaged to explore the processing pattern related to problem triggers. Korpal and Stachowiak-Szymczak (2018) aimed to explore the number processing with the context. They prepared slides containing the numbers of the speech and important content information displayed in bullet points. Mean fixation duration was selected as a reliable indicator for comparing the cognitive effort involved in processing numbers and the context. Longer fixation duration was found in processing numbers which indicates both professionals and trainees spent more effort on numbers than on the context. Their data analysis about the interpreting performance also proved a positive correlation between processing numbers and their context. Their findings did not corroborate with the spillover effect (Gile, 2009), although number processing requires higher cognitive effort, it did not affect the accuracy of processing their context. Following this line, they made a few more attempts to investigate the number processing patterns in simultaneous interpreting. Stachowiak-Szymczak and Korpal (2019) switched their focus to comparing the number processing pattern between professional and trainee groups. Participants were guided to simultaneously interpret a speech with visual information from slides containing numbers. Their findings suggested two processing patterns: 1) Although professional interpreters were observed to have more fixations on numbers than trainees, the fixation was relatively shorter thus leading to shorter total gaze time; 2) In general, trainee groups were found to have longer fixation and longer gaze time which indicates trainees devote more effort in processing numbers. Korpal and Stachowiak-Szymczak (2020) added another problem trigger delivery rate to number processing. Participants were presented with slides containing key points such as names and numbers, and they were tested in two conditions: fast and slow. They observed a higher fixation count in number processing at a fast delivery rate, this indicated an increased cognitive effort involved. However, no significant effect of delivery rate on the percentage of gaze time spent on numbers was found. They explained this might be because participants allocated similar time to process numbers and other information (i.e., context and figures). Chmiel et al. (2020) manipulated three types of information (numbers, proper names and control words) to test the incongruences of audio and visual input in simultaneous interpreting. Their results reported higher accuracy in the congruent condition than incongruent condition, and numbers enjoyed the highest accuracy among these three types of information.

As demonstrated in the literature reviewed in this section, the cognitive processing of live captioning is worth investigating from the following three aspects: 1) Multimodal processing in simultaneous interpreting has attracted increasing attention. Simultaneous interpreters are generally believed to be multitasking (Chmiel et al., 2020; Gile, 2009; Lambert, 2004) since they have to process information from multiple channels. Previous studies have explored the cognitive effort related to processing information, in particular, nonverbal channels (i.e., Stachowiak-Szymczak and Korpal, 2019) and how it might affect interpreting performance (i.e., Lambert, 2004). The extra layer input of live transcripts can be seen as a new form of multimodal resources which might pose additional cognitive challenges; 2) the (in)congruence between visual and verbal information has also been examined. Incongruent visual information has been proven to decrease interpreting performance (Korpal and Stachowiak, 2015) and cause higher cognitive load (Chmiel et al., 2020). From this perspective, live captioning enabled on the Zoom platform is worth investigating since the live captioning technology at this stage is not error-free. Due to the limitation of voice recognition, the information presented by live captioning

(such as numbers) can be congruent while other types of information might be incongruent; 3) Problem triggers such as numbers have been investigated in different forms, for instance, presented on slides (Stachowiak-Szymczak and Korpal, 2019), transcriptions (Seeber et al., 2020) and in sentences (Chmiel et al., 2020). Scrolling live captioning on the screen is a new form which entails all information (including numbers).

To this date, no study has been conducted to explore the cognitive challenges brought by live captioning. Little has been known about cognitive effort related to the incongruent nature of live captioning, as well as how interpreters will process the information presented in live captioning, for instance, numbers and proper names. Against this background, the current study aims to fill in the research gap and contribute some empirical data in exploring the cognitive processing of live captioning in simultaneous interpreting.

3. Research questions

The aim of this study is to investigate live captioning in simultaneous interpreting on Zoom the remote meeting platform. The current study utilized eye-tracking technology to examine the cognitive processing of live captions. In order to address the research objectives, the current study aims to answer the following research questions:

RQ 1: How do the interpreters process the additional layer of visual information from live captioning?
RQ 2: Which types of information segments tax more cognitive resources in interpreting with live captioning?
RQ 3: Is there a significant difference in interpreting accuracy between interpreting with live captioning and interpreting without live captioning?

RQ 1 looked at how the independent variable of information type affects the dependent variables of run count and percentage of dwell time. RQ 2 delves into understanding how variations in the two independent variables of information type and task impact fixation count per second and average fixation duration. Finally, RQ 3 focuses on the influence of information type and task conditions on interpreting performance.

4. Experimental design

4.1. Participants

The selection of participants for this study was carefully executed to ensure a high level of homogeneity.13 participants were recruited from one of the word-level programs in professional interpreting training. Despite its modest size, the inclusion of 13 participants from a globally renowned program in professional interpreting training ensures the necessary quality and homogeneity for a representative cohort.

All participants share Chinese as their first language (L1) and English as their second language (L2). They are registered in the same course and have undergone a comparable amount of interpreting practice within the classroom setting. Prior to this experiment, all participants have completed a minimum of six months of specialized training in simultaneous interpreting. This training duration establishes a foundational level of proficiency and competence in simultaneous interpreting. All the participants had a normal or corrected-to-normal vision.

4.2. Design

This study is an eye-tracking experiment of the E-C simultaneous interpreting, eye movement data and the interpreting recordings were recorded. The experiment adopted a 2 × 2 within-subject design. The independent variables include information type (with numbers and proper names vs without numbers and proper names) and task conditions (with live captioning on vs with live captioning off). The following dependent variables were selected:

- fixation count per second: fixation count refers to the number of fixations in a given area (Holmqvist et al., 2015, p. 412). It is a frequently selected indicator to measure cognitive effort in interpreting studies (Korpal and Stachowiak-Szymczak, 2020; Stachowiak-Szymczak and Korpal, 2019). Given the interest periods of the two speeches were not fully matched in time lengths, fixation count per second was calculated for comparing the cognitive load in those interest periods.
- average fixation duration: it refers to "the sum of the duration of the durations of all fixations divided by the number of fixations" (Chen et al., 2021). It is a widely adopted indicator to investigate cognitive processing in interpreting studies (Ho, 2021; Korpal and Stachowiak-Szymczak, 2018; Stachowiak-Szymczak and Korpal, 2019; Su, 2020). In general, longer fixation duration is often "associated with deeper and more effortful cognitive processing (Holmqvist et al., 2015, p. 515).
- run count: it refers to the total number of fixations runs in the trial. It can be used to calculate the times the number of an AOI was entered and left (Eyelink, 2018).
- the percentage of dwell time: it refers to the percentage of trial time spent on the interest areas (Eyelink, 2018). Dwell time, also known as gaze time, is a commonly selected eye movement indicator for the investigation of visual attention (Korpal and Stachowiak-Szymczak, 2020; Seeber, 2012).
- interpreting performance: the interpreting performance was assessed based on the errors observed in two task conditions. The purpose of utilizing an error-based analysis is to compromise the subjective caused by different makers recruited for interpreting performance assessment. This analysis was based on the classification of error types proposed by (Barik, 1971) in simultaneous interpreting. For instance, omissions refer to "items present in the original version which are left out of the translation" and additions refer to "materials which are added outright the text".

To obtain the research objectives, the following hypotheses are formulated:

1) When interpreting with live captioning, participants will:
   a) have less run count between face and live captioning areas when processing information with numbers and proper names.
   b) allocate more attention to the live captioning area when numbers and proper names appear on the screen.
2) When interpreting with live captioning, information segments containing numbers and proper names will be more demanding than those without. Participants will:
   a) have more fixations on information with numbers and proper names.
   b) generate longer average fixation duration on information with numbers and proper names than those without.
3) Interpreting accuracy on numbers and names will be higher in interpreting with live captioning.

4.3. Materials

The experiment stimuli were a speech adapted from an article in *The Economist*. The rationale behind the selection of an article from The Economist was to test the effectiveness of live captioning by incorporating moderately difficult material. The adaptation of the speech materials was primarily content based. The focus of the adaption lies in the oral rendition of the speech, aiming to faithfully convey the message of this article instead of exclusively adhering to its written form. The present study invited a native English speaker to refine the article into a genuine oral speech, and later video recorded by the same speaker on

the Zoom platform. Half of the speech was recorded with live captioning on, and the other half was recorded with live captioning off.

16 periods of interest (POIs) were created in these two speeches. 8 periods of interest were created in the task with live captioning on (4 POIs containing numbers and proper names and 4 POIs without numbers and proper names), 8 POIs were created in the control task with live captioning off (4 POIs containing numbers and proper names and 4 POIs without numbers and proper names). To minimize the impact of source-text-related features and to create a well-controlled experimental setting, the present study adopted the method of the Flesch Reading Ease formula (Liu and Chiu, 2009). Readability scores, such as the Flesch Reading Ease Score, were calculated by considering factors such as sentence length and word difficulty.16 POIs were comparable in terms of difficulty level, average word per sentence and average sentence duration. The delivery rate for POIs with and without live captioning was 113 wpm and 112 wpm. Detailed information is provided in Table 1.

4.4. Procedures and apparatus

The empirical eye-tracking experiment was meticulously conducted within a controlled laboratory setting, specifically designed to minimize extraneous influences. The laboratory environment was acoustically isolated to ensure a sound-proof environment, and illumination was provided by consistent artificial lighting conditions. Prior to the experiment, participants were provided with a comprehensive explanation of the study's objectives, specifically focusing on investigating the processing of live captioning in simultaneous interpreting on the Zoom platform. Participants were guided to read the information sheet about the experiment and signed the consent forms.

Following the consent process, participants were instructed to position themselves in front of a high-resolution display screen measuring 53.1 cm in width and 29.7 cm in height. The eye-tracker employed for data collection was the Eye Link 1000 Plus, strategically positioned at an approximate distance of 92 cm from the participants to ensure optimal tracking accuracy.

During the experimental sessions, participants were required to engage in simultaneous interpreting tasks on the Zoom platform, under two distinct conditions: one involving the presence of live captioning, and the other with live captioning disabled. To ensure accurate eye-tracking measurements, a meticulous 9-point calibration procedure was conducted before each task. The entire experimental session, encompassing participant preparation, calibration procedures, and the execution of the simultaneous interpreting tasks, lasted approximately 30 min for each participant.

The original eye-tracking experiment was programmed by Experiment Builder 2.3.38. Participants' eye movements were tracked with a sampling rate of 500 Hz. The eye movement data were imported into the Eye link Data Viewer 4.3.1 for data analysis. Their interpretation was recorded by the OSX audio drive in the macOS version (11.5.2) throughout the experiment. IBM SPSS Statistics 26 was adopted for statistical analysis.

Table 1
Details of the POIs.

Condition	Reading Ease score		Average word per sentence		Average sentence duration	
	M	SD	M	SD	M (sec)	SD (sec)
1	53.3	1.3	13.7	0.2	25.8	0.2
2	52.5	1.2	13.3	0.5	25.8	0.2
3	52.3	1.1	13.3	0.5	25.4	0.4
4	53.2	1.3	13.2	0.2	25.2	0.4
Total	52.5	1.2	13.4	0.4	25.6	0.4

Condition 1: Live captioning on (with number and proper names).
Condition 2: Live captioning on (without number and proper names).
Condition 3: Live captioning off (with number and proper names).
Condition 4: Live captioning off (without number and proper names).

4.5. Data collection

4.5.1. Eye tracking data

Eye movement data were collected from a cohort consisting of 13 participants, and subsequent analysis involved applying rigorous filtering thresholds. The objective of this filtering process was to ensure the data quality by excluding extreme values that could potentially distort the overall fixation patterns. Such extreme values, which may arise from individual differences or technical recording issues, contain fixations that were either excessively short or abnormally long. Fixations lasting less than 80 ms or exceeding 1200 ms were removed from the final dataset, following established recommendations from prior research (Drieghe et al., 2008). Additionally, in line with the empirical evidence presented by Pavlović and Jensen (2009) and Ma and Li (2021), a cut-off point of 200 ms was employed as the average fixation duration threshold. Due to concerns that shorter fixations could potentially indicate errors or recording instability (Ma, 2021), two participants in this study were excluded from the analysis as approximately half of their fixations fell below the 200 ms threshold. As a result, data from 11 participants were employed for the analysis of eye-tracking data.

The present study aims to investigate the cognitive processing pattern of live captioning. To explore which types of information segments tax more cognitive resources in interpreting with captioning, 16 POIs were created for the whole experiment (8 POIs for each task condition).

4 POIs were information segments without numbers and proper names, and the other 4 POIs were information segments with numbers and proper names. To assess the cognitive resources, fixation count per second and average fixation duration were collected for further analysis. To examine how interpreters process the additional layer of visual information from live captioning, the speaker's face and live caption areas were drawn as two Areas of Interest (AOIs). Run count and percentage of dwell time were collected to examine the processing pattern. In addition, time course analysis graphs were drawn to present a detailed temporal attention distribution in real-time.

4.5.2. Interpreting performance data

To assess the interpreting quality in two conditions, the transcription of the interpretation and recording were prepared for an experienced interpreter trainer who has 6 years of teaching experience. In order to reduce the subjectivity of the performance assessment, this study adopted an error-based analysis with a focus on the assessment of POIs containing numbers and proper names in two conditions. The criteria of the error analysis were drawn from the classification of error types proposed by Barik (1971). Barik (1971) suggested that during simultaneous interpreting, interpreters might "depart from the original version". Three types of errors were generalized namely omission, addition and substitution. Omission refers to the items in the source speech which were left out in the interpretation of the interpreter. Addition in general indicates items which are completely added to the original version by the interpreter. Substitution and errors refer to items that were replaced by the interpreter for things expressed by the speaker.

4.5.3. Statistical analysis

All the data were collected and submitted to SPSS Statistic 26 to test the distribution of normality. Shapiro-Wilk test was conducted to test the normality. In terms of the data conformed to the normality distribution, paired *t*-test for within-subject comparison of the interpreting performance. Two-way repeated measures ANOVAs were administered to compare the eye movement indicators. If the data were not conformed to normality distribution, non-parametric tests would be administered.

5. Results

5.1. Processing pattern of live captioning in interpreting

To investigate how interpreters process the additional layer of visual information from live captioning, the speaker's face and live captioning areas were drawn as two Areas of Interest (AOIs). Run count and percentage of dwell time were collected for further analysis.

The data of run count were not conformed to normality distribution. The Wilcoxon signed-rank tests were performed to compare the run count in processing POIs with numbers and proper names and POIs without numbers and proper names. No significant differences ($Z = -1.531$, $p = 0.13$) were observed between POIs with numbers and proper names ($M = 4.27$, $SD = 4.95$) and POIs without numbers and proper numbers ($M = 4.95$, $SD = 3.24$). See Table 2 for descriptive statistics.

Although no statistical difference was observed, POIs with numbers and proper names resulted in fewer attention shifts. Constant attention shifts between face and live captioning were observed in both types of POIs.

To investigate how interpreters process the live captioning when there were numbers and proper names, the percentage of dwell time offers a holistic view of attention distribution. The percentage of dwell time was not conformed to normality distribution. The Wilcoxon signed-rank tests were administered to explore the attention allocated to the speaker's face and live captioning area for POIs with numbers and proper names. Significant differences were observed in the amount of attention allocated to face and live captioning areas ($Z = -5.712$, $p < 0.001$). The amount of attention devoted to live captioning areas ($M = 0.76$, $SD = 0.2$) was significantly higher than face areas ($M = 0.21$, $SD = 0.22$). See Table 3.

To probe into details, time course analysis graphs were drawn to plot the attention distribution when processing numbers and proper names in real-time. The time interval of −500ms–500ms signifies the period during which numbers and proper names are displayed on the screen. The negative −2000ms and positive 2000ms time periods encompass the time before and after this interval, respectively. The selected time range of −2000ms–2000ms was determined as the maximum duration suitable for the experimental design of the present study. Expanding the time period beyond this range carries the potential for overlap with other temporal regions that also contain numbers and proper names. Given the frequent occurrence of numbers and proper names presented on the screen within POIs, it is crucial to maintain a clear distinction and avoid any confounding effects between the target stimuli and unrelated temporal intervals.

Fig. 1 shows the temporal change of attention distribution in POIs with numbers and proper names.

As demonstrated above in Fig. 1, during the time period from −500ms to 500ms where there is a high density of numbers and proper names, the amount of attention allocated to the live captioning area reached the maximum amount. This is indicated by the nearly 100% sample count while the 0% sample count was found for the speakers' face area.

5.2. Cognitive effort devoted to different types of information

It was expected that when processing different types of information, participants would devote more cognitive effort to processing information with numbers and proper names, as characterized by more fixation count and average fixation duration.

Table 2
Mean (SD) values for run count.

Condition	M	SD	SEM
POIs with numbers	4.27	4.95	0.65
POIs without numbers	4.95	3.24	0.49

Table 3
Mean (SD) values for the percentage of dwell time.

AOIs	M	SD	SEM
Face	0.21	0.22	0.33
Live captioning	0.76	0.2	0.31

When analysing the eye-tracking data, fixation count per second and average fixation duration were collected for statistical analysis. 8 POIs were created for the task condition with live captioning (4 POIs without numbers and proper names and 4 POIs with numbers and proper names), and the same applied to the task condition without live captioning. Fixation count per second and average fixation duration was collected for the total 16 POIs in both task conditions.

5.2.1. Fixation count per second

The fixation count per second conformed to normality distribution according to the Shapiro-Wilk test. Two-way repeated measures ANOVA was adopted. Descriptive statistics is presented in Table 4.

In terms of the fixation count per second, we observed a main effect of live captioning ($F(1,43) = 179.27$, $p < 0.001$). The fixation count per second was significantly higher in Condition 1 than in Condition 2 of all POIs. In terms of processing POIs with numbers and proper names, statistical differences were found ($p < 0.001$) in the fixation count per second between Condition 1 ($M = 3.16$, $SD = 0.4$) and Condition 2 ($M = 1.95$, $SD = 0.51$). With regard to processing POIs without numbers and proper names, statistical differences were also found ($p < 0.001$) in the fixation count per second between Condition 1 ($M = 3.02$, $SD = 0.43$) and Condition 2 ($M = 2.13$, $SD = 0.49$).

However, there was no significant main effect of numbers and proper names ($F(1,43) = 0.209$, $p = 0.649$). The fixation count per second was significantly higher in processing the POIs with numbers and proper names ($M = 3.16$, $SD = 0.4$) than POIs without ($M = 3.02$, $SD = 0.43$). Statistical differences were observed ($p = 0.002$). However, Condition 2 revealed the opposite pattern. The fixation count per second was significantly higher in processing the POIs without numbers and proper names ($M = 2.13$, $SD = 0.49$) than POIs with numbers and proper names ($M = 1.95$, $SD = 0.51$). The statistical difference was proved ($p < 0.001$).

5.2.2. Average fixation duration

According to the Shapiro-Wilk test, the average fixation duration for both conditions was not conformed to normality distribution. The Wilcoxon signed-rank tests were administered to compare the average fixation duration of different POIs in both conditions. No significant differences were found in the average fixation duration in processing POIs with numbers and proper names ($M = 267.32$, $SD = 41.37$) than POIs without ($M = 274.55$, $SD = 53.47$) in Condition 1($Z = -1.109$, $p = 0.268$). No significant differences were found in the average fixation duration in processing POIs with numbers and proper names ($M = 436.21$, $SD = 133.68$) than POIs without ($M = 415.12$, $SD = 113.14$) in Condition 2 ($Z = -1.599$, $p = 0.11$). See Table 5.

5.3. Comparison of interpreting accuracy between the two conditions

An error-based analysis method was adopted to evaluate the interpretation with a focus on the interpreting accuracy of numbers and proper names. One of our research questions is to examine if there is a significant difference in interpreting accuracy between interpreting with live captioning and interpreting without live captioning. Our hypothesis is that interpreters have better accuracy in numbers and proper names with live captioning on.

Error rates in both task conditions were identified by an experienced interpreting trainer. The Shapiro-Wilk test was administered to check the data normality. The data conformed to the normality distribution ($p = 0.652$). Paired t-tests were administered to compare the error

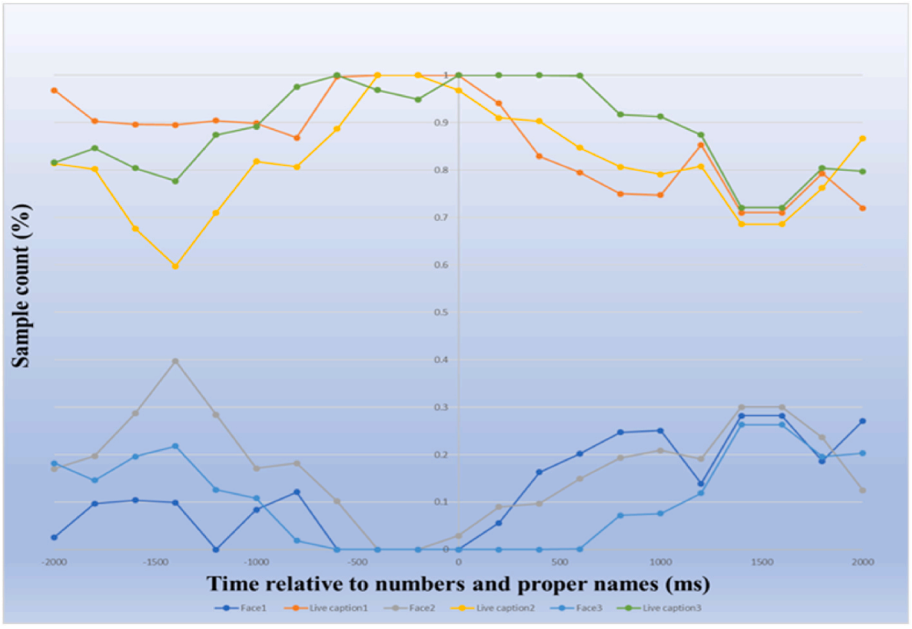

Fig. 1. Time (bin) course graphs showing sample count (%) on the face and live caption.

Table 4
Mean (SD) values of fixation count per second.

Condition	M	SD	SEM
1 (POIs with numbers)	3.16	0.4	0.61
1 (POIs without numbers)	3.02	0.43	0.65
2 (POIs with numbers)	1.95	0.51	0.52
2 (POIs without numbers)	2.13	0.49	0.74

Condition 1: interpreting with live captioning; **Condition 2**: interpreting without live captioning.

Table 5
Mean (SD) values of average fixation duration (ms).

Condition	M	SD	SEM
1 (POIs with numbers)	267.32	41.37	6.24
1 (POIs without numbers)	274.55	53.47	8.06
2 (POIs with numbers)	436.21	133.68	20.15
2 (POIs without numbers)	415.12	113.14	17.06

Condition 1: interpreting with live captioning; **Condition 2**: interpreting without live captioning.

frequencies of POIs containing numbers and proper names in both task conditions. A significant difference was observed in both conditions ($t = -7.396$, $p < 0.001$). The error rate for POIs with live captioning on was 23.2% (SD = 0.12), and for POIs with live captioning off was 53.6% (SD = 0.16). Descriptive statistics for error rates are presented in Table 6. Condition 1 represents the task with live captioning, and Condition 2 refers to the task without live captioning.

According to the statistics, the error rates decreased by 30% in interpreting POIs with numbers and proper names in Condition 1. Among the error types, omission is the most identified error type in both conditions. For example, one participant interpreted "By 2015, when Apple launched its first watch" into "当苹果发布了它的第一款手表时". The number 2015 was completely left out of the source speech. The omission rates were 70% (SD = 0.27) and 72% (SD = 0.094) for Condition 1 and Condition 2 respectively. No statistical difference was observed in both conditions ($t = -1.3$, $p = 0.8$).

Table 6
Mean (SD) values of error rates in both task conditions.

Condition	Mean	SD	SEM
1	0.232	0.12	0.35
2	0.536	0.16	0.47

6. Discussion

As seen from the data presented in Section 5.1, hypothesis 1a was rejected. As indicated by the run count data, there was no significant difference found between POIs containing numbers and proper names and those without. Participants did not have less shift of attention between face and live captioning areas when there were numbers and proper names. This might be explained by the in(congruences) of the live captioning. The initial expectation of this study was that live captioning would largely offer accurate numbers and proper names. Surprisingly, the results of our experiment demonstrated a full congruence of numbers and proper names between live captioning and the source text. This might be attributed to the high accuracy rate of Zoom's voice recognition technology. However, there still exists the incongruences of other types of information that might lead interpreters to shift their attention away to reduce their cognitive load. This partly sided with the findings from Chmiel et al. (2020) that incongruences between audio and visual channels are likely to increase the cognitive load for simultaneous interpreters. In terms of hypothesis 1b, a higher percentage of dwell time in the live captioning area in POIs with numbers indicating that most of the attention was allocated to the live captioning area. Also, as seen from the time bin course analysis graphs, it is obvious that the attention on the live transcript peaked when numbers and proper names appeared on the screen. This also confirms the findings that interpreters are actively searching for visual information that might be complementary to the audio input (Korpal and Stachowiak-Szymczak, 2020; Seeber, 2012).

Based on the data presented in Section 5.2, fixation count per second was observed to be higher in POIs containing numbers and proper names than in those without. Hypothesis 2a was confirmed. A higher fixation count points to a higher cognitive effort involved in processing information segments containing numbers and proper names. As mentioned, live captioning is dynamic scrolling on the screen and not error-free, a higher fixation count might also indicate that interpreters might

mobilise more cognitive resources to actively search for the numbers and proper names presented in live captioning. This is in line with the previous finding from previous studies (Chmiel et al., 2020; Korpal and Stachowiak-Szymczak, 2020). Additionally, higher fixation counts were found in interpreting with live captioning than without live captioning in general. This indicates visually processing the extra layer of live captioning does increase the cognitive effort and confirmed the findings from (Chmiel et al., 2020; Seeber et al., 2020). However, hypothesis 2b was rejected. There was no significant difference observed in the average fixation duration between POIs containing numbers and proper names and those without. This contradicts the finding from Korpal and Stachowiak-Szymczak (2018) that a longer average fixation duration was observed on processing numbers. It can be explained that with live captioning on Zoom, the text is scrolling on the screen and it is not static as in the slides used by Korpal and Stachowiak-Szymczak (2018). The fixations generated were relatively shorter on the screen according to the constant change of live captioning, hence, leading to no significant difference in the average fixation duration.

Finally, interpreting performance data from Section 5.3 proved that the presence of live captioning does affect the interpreting accuracy for numbers and proper names. The error rates decreased by 30% in interpreting with live captioning, the third hypothesis is thus corroborated. Gile (2009) pointed out problem triggers such as numbers and proper names can be challenging for simultaneous interpreting, our finding pointed to a facilitative role of live captioning. This is in line with the finding from the previous experimental studies on various types of visual materials in number rendition. For instance, visual materials based on automatic speech recognition technology help to reduce the error rate of numbers dropped significantly (Defrancq and Fantinuoli, 2021; Desmet et al., 2018); visual access to slides containing numbers also improved the interpreting performance (Korpal and Stachowiak-Szymczak, 2020; Stachowiak-Szymczak and Korpal, 2019).

7. Conclusion

The present study explored the cognitive processing pattern of live captioning in simultaneous interpreting on Zoom. As live captioning contains various types of information from the source speech, this study was designed with a focus on the numbers and proper names which are considered problem triggers. As seen from the eye movement data, when processing the live captioning on the screen, interpreting trainees tend to actively search for the presence of numbers and proper names on the screen. Their attention on the live captioning area peaked during the period when there is a high density of numbers and proper names. However, the presence of numbers and proper names did not generate fewer attention shifts between the speaker's face and the live captioning area. There was not a significant difference in the attention shifts between information segments containing numbers and proper names and those without. This might indicate that although interpreting trainees were actively searching for numbers and proper names, they tend to shift their attention away when encountering incongruent information in the live captioning area to avoid exceeding their processing capacity. In addition, interpreting trainees devoted more cognitive effort to processing information segments containing numbers and proper names compared to other types of information. As seen from the interpreting performance data, the live captioning offered on Zoom does affect the interpreting accuracy of numbers and proper names. The error-based analysis results pointed to a facilitative role of live captioning as indicated by fewer error rates.

We hope that our findings can offer some tentative perspectives for cognitive processing studies on live captioning in simultaneous interpreting. Due to the rapid development of voice recognition technology, live captioning offered on the Zoom platform has posed new challenges not only for professionals but also for interpreting trainers. We believe that our findings offer some insights to both professionals and interpreting trainers about the potential benefits of live captioning and how to exploit it skillfully in real-life tasks.

The study is admittedly limited in several ways: 1) One potential limitation of the experiment pertains to the use of the remote Zoom platform as the medium for conducting the study. The recorded experimental stimuli were presented within the confines of the Zoom interface. Specifically, the live captioning area occupied a relatively small portion of the bottom screen, resulting in limited visibility and potential constraints for detailed analysis of eye movement data at the word level. The close proximity of word characters in the live captioning display further complicated the accurate extraction and interpretation of fine-grained eye movement patterns. Consequently, the ability to explore in-depth cognitive processing related to living captioning in this experimental setting was restricted. To address this limitation, future studies could consider alternative experimental setups that offer a more comprehensive investigation of this field; 2) The present study focused on the specific Chinese-English language pair, the applicability of the findings to other language pairs might be limited due to linguistic and cultural factors. Future studies incorporating a broader range of language pairs are crucial in obtaining a more comprehensive understanding of this area; 3) This study did not recruit professional interpreters. Since the pandemic, professional simultaneous interpreters have already been exposed to the function of live captioning on Zoom for a while. Given this and their professionalism, they might present different processing patterns compared to the trainees. Additionally, it is important to acknowledge the limitation of this study's small sample size. The small number of interpreting trainees in highly professional programs prohibits the attainment of a large sample size comparable to that of translation or language studies. Future research is needed to conduct larger studies encompassing a more diverse range of participants.

Note

In this co-authored paper, the first author did data-analysis and drafting of the manuscript; the second author conceptualised and designed the research and revised the draft critically for intellectual content.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lu Yuan reports writing assistance was provided by University of Leeds. Lu Yuan reports a relationship with University of Leeds that includes: non-financial support.

References

Ahrens, B., 2017. Interpretation and cognition. In: The Handbook of Translation and Cognition. John Wiley & Sons, Ltd, pp. 445–460. https://doi.org/10.1002/9781119241485.ch24.

Amos, R., Seeber, K., Pickering, M., 2022. Prediction during simultaneous interpreting: evidence from the visual-world paradigm. Cognition 220. https://doi.org/10.1016/j.cognition.2021.104987.

Barik, H.C., 1971. A description of various types of omissions, additions and errors of translation encountered in simultaneous interpretation. Meta: Journal des traducteurs 16 (4), 199. https://doi.org/10.7202/001972ar.

Bühler, H., 1985. Conference interpreting: a multichannel communication phenomenon. Meta 30 (1), 49–54.

Chen, S., Kruger, J.-L., Doherty, S., 2021. Reading patterns and cognitive processing in an eye-tracking study of note-reading in consecutive interpreting. Interpreting 23 (1), 76–102. https://doi.org/10.1075/intp.00050.che.

Chmiel, A., Janikowski, P., Lijewska, A., 2020. Multimodal processing in simultaneous interpreting with text: interpreters focus more on the visual than the auditory modality. Target. International Journal of Translation Studies 32 (1), 37–58. https://doi.org/10.1075/target.18157.chm.

Defrancq, B., Fantinuoli, C., 2021. Automatic speech recognition in the booth Assessment of system performance, interpreters' performances and interactions in the context of numbers. Target-International Journal of Translation Studies 33 (1), 73–102. https://doi.org/10.1075/target.19166.def.

Desmet, B., Vandierendonck, M., Defrancq, B., 2018. Simultaneous interpretation of numbers and the impact of technological support. In: Interpreting and Technology. Language Science Press, pp. 13–27.
Drieghe, D., Pollatsek, A., Staub, A., Rayner, K., 2008. The word grouping hypothesis and eye movements during reading. J. Exp. Psychol. Learn. Mem. Cognit. 34 (6), 1552.
Eyelink, 2018. EyeLink Data Viewer User Manual. Canada. SR Research Ltd version 3.2.1.
Galvão, E.Z., 2013. Hand gestures and speech production in the booth: Do simultaneous interpreters imitate the speaker? In Carapinha C and Santos IA (eds), Estudos de linguística. Coimbra: Imprensa da Universidade de Coimbra Vol. II,, 115–130.
Gile, D., 2009. Basic Concepts and Models for Interpreter and Translator Training. John Benjamins Pub. Co.
Ho, C.-E., 2021. What does professional experience have to offer?: an eyetracking study of sight interpreting/translation behaviour. Translation, Cognition & Behavior 4 (1), 47–73. https://doi.org/10.1075/tcb.00047.ho.
Holmqvist, Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2015. Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press.
Hyönä, J., Tommola, J., Alaja, A.-M., 1995. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. The Quarterly Journal of Experimental Psychology Section A 48 (3), 598–612. https://doi.org/10.1080/14640749508401407.
Just, M.A., Carpenter, P.A., 1980. A theory of reading: from eye fixations to comprehension. Psychol. Rev. 87 (4), 329–354.
Korpal, P., Stachowiak, K., 2013. In: Numerical Data Processing in Simultaneous Interpreting: Does Visual Input Facilitate the Process? Poster Presented at the 11th Symposium of Psycholinguistics, Tenerife, 20–23 March 2013.
Korpal, P., Stachowiak, K., 2015. The visual or the aural: which modality is dominant in simultaneous interpreting?. In: Conference Paper Presented at the ICEAL (International Conference on Eyetracking and Applied Linguistics), Warszawa, September 22.
Korpal, P., Stachowiak-Szymczak, K., 2018. The whole picture: processing of numbers and their context in simultaneous interpreting. Poznań Stud. Contemp. Linguistics 54 (3), 335–354. https://doi.org/10.1515/psicl-2018-0013.
Korpal, P., Stachowiak-Szymczak, K., 2020. Combined problem triggers in simultaneous interpreting: exploring the effect of delivery rate on processing and rendering numbers. Perspectives 28 (1), 126–143. https://doi.org/10.1080/0907676X.2019.1628285.
Kress, G., Van Leeuwen, T., 2001. Multimodal Discourse: the Modes and Media of Contemporary Communication. Arnold Publishers, London.
Kruger, J.-L., Steyn, F., 2014. Subtitles and eye tracking: reading and performance. Read. Res. Q. 49 (1), 105–120. https://doi.org/10.1002/rrq.59.
Kuang, H., Zheng, B., 2022. Note-taking effort in video remote interpreting: effects of source speech difficulty and interpreter work experience. Perspectives 1–21. https://doi.org/10.1080/0907676X.2022.2053730.
Lambert, S., 2004. Shared attention during sight translation, sight interpretation and simultaneous interpretation. Meta 49 (2), 294–306. https://doi.org/10.7202/009352ar.
Liu, M., Chiu, Y.H., 2009. Assessing source material difficulty for consecutive interpreting: quantifiable measures and holistic judgment. Interpreting 11 (2), 244–266.
Ma, X., Li, D., 2021. A cognitive investigation of 'chunking'and 'reordering'for coping with word-order asymmetry in English-to-Chinese sight translation: evidence from an eye-tracking study. Interpreting 23 (2), 192–221.
Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., Tessendorf, S., 2013. Body-language-Communication. An international handbook on multimodality in human interaction 1 (1), 131–232.
Pavlović, N., Jensen, K., 2009. Eye tracking translation directionality. Translation research projects 2, 93–109.
Pöchhacker, F., 2021. Multimodality in interpreting. In: Handbook of Translation Studies. John Benjamins Publishing Company, Amsterdam, pp. 152–158.
Poyatos, Fernando, 1983. Language and nonverbal systems in the structure of face-to-face interaction. Lang. Commun. 3 (2), 129–140.
Rodrigues, Galhano, Isabel, 2007. Body in interpretation: nonverbal communication of speaker and interpreter and its relation to words and prosody. In: Schmitt, P.A., Jüngst, H. (Eds.), Translationsqualität. Peter Lang, Frankfurt, pp. 739–753.
Seeber, K., 2012. Multimodal input in simultaneous interpreting: an eye-tracking experiment. In: In*Proceedings of the 1st International Conference TRANSLATA, Translation & Interpreting Research: Yesterday-Today-Tomorrow*. Peter Lang.
Seeber, K.G., 2017. Multimodal processing in simultaneous interpreting. In: Schwieter, J.W., Ferreira, A. (Eds.), The Handbook of Translation and Cognition, first ed. Wiley, pp. 461–475. https://doi.org/10.1002/9781119241485.ch25.
Seeber, K.G., Keller, L., Hervais-Adelman, A., 2020. When the ear leads the eye – the use of text during simultaneous interpretation. Language, Cognition and Neuroscience 35 (10), 1480–1494. https://doi.org/10.1080/23273798.2020.1799045.
Serbert, S., 2019. Visuelle Informationen Beim Simultandolmetschen: Eine Eyetracking-Studie, vol. 47. Frank & Timme GmbH.
Shlesinger, Miriam, 1994. Intonation in the production and perception of simultaneous interpretation. In: Lambert, S., Moser-Mercer, B. (Eds.), Bridging the Gap: Empirical Research in Simultaneous Interpretation. John Benjamins, Amsterdam, pp. 225–236.
Stachowiak, K., 2017. Eye Movements and Gestures as Correlates of Language Processing in Consecutive and Simultaneous Interpreting. PhD Dissertation. Adam Mickiewicz University, Poznań.
Stachowiak-Szymczak, K., Korpal, P., 2019. Interpreting accuracy and visual processing of numbers in professional and student interpreters: an eye-tracking study. Across Lang. Cult. 20 (2), 235–251. https://doi.org/10.1556/084.2019.20.2.5.
Su, W., 2020. Eye-Tracking Processes and Styles in Sight Translation. Springer Singapore. https://doi.org/10.1007/978-981-15-5675-3.
Vranjes, J., Brône, G., 2021. Interpreters as laminated speakers: gaze and gesture as interpersonal deixis in consecutive dialogue interpreting. J. Pragmat. 181, 83–99. https://doi.org/10.1016/j.pragma.2021.05.008.
Wang, B., 2018. Exploring Approaches to Interpreting Studies: from semiotic perspectives to multimodal analysis. Chinese Semiotic Studies 14 (2), 149–161. https://doi.org/10.1515/css-2018-0010.
Zou, L., Carl, M., Feng, J., 2022. Patterns of attention and quality in English-Chinese simultaneous interpreting with text. International Journal of Chinese and English Translation & Interpreting. https://doi.org/10.56395/ijceti.v2i2.50.