ARTICLE

https://doi.org/10.1057/s41599-025-04492-w

OPEN

Check for updates

The function of ASR-generated live transcription in simultaneous interpreting: trainee interpreters' perceptions from post-task interviews

Xiaoman Wang¹, Binhua Wang^{2⊠} & Lu Yuan²

This study explores whether live transcription generated with the technology of automaticspeech-recognition (ASR) can be used to facilitate simultaneous interpreting. This article reports an analysis of trainee interpreters' perceptions based on post-task structured interviews after an eve-tracked interpreting task without live transcription in the first half and with live transcription in the second half, which was done by a group of trainee interpreters from a postgraduate professional interpreting programme. The interviews were analysed in triangulation with the eye-tracking data about their interpreting behaviours. The results show that most participants perceived live transcription beneficial, with data indicating improved performance and lowered error rates in terms of terminologies, numbers, and proper names. It is also found that while some interpreters reported that they can adeptly manage multimodal inputs, others reported challenges in optimizing their focus of attention when live transcription was provided. The overall interference score in interpreting with live transcription spikes from 9 to 13.2, suggesting fluctuating cognitive demand. Eye-tracking data further corroborate these attentional dynamics, echoing participants' self-reported behaviours. The study points to the need for training programmes to equip interpreters with capabilities to utilize technological tools such as live transcription, ensuring optimal attention management and overall performance.

Introduction

n an era where the remote conference has replaced the conference room, and the screen stands in for the stage, interpreters are navigating a significant "technological turn" (Fantinuoli 2018, p.1). Some remote conferencing platforms such as Zoom Meeting stands out for its dual functionality: it serves as a platform for remote interpreting and offers simultaneous interpreting with live transcription, a feature with potential for computer-assisted interpreting. This multifaceted role of Zoom raises compelling questions about the impact of digital aids, like live transcription, on interpreting performance. Simultaneous interpreting has long been a complex cognitive exercise, demanding not just linguistic skills but also a delicate balancing act of attention and comprehension (Gerver 1969; Gile 1985, 1997, 2009; Kirchhoff 1976). Traditionally, interpreters have relied on auditory stimuli, but the digital age introduces new layers to this process. The advent of Zoom and its features like live transcription adds an intricate web of visual cues to the auditory mix. While live transcription appears to be a boon, offering an additional layer of information, its cognitive implications are still not fully understood.

This study aims to explore whether live transcription generated with the technology of automatic-speech-recognition (ASR) can be used to facilitate simultaneous interpreting. Structured interviews are conducted about trainee interpreters' perceptions with a group of trainee interpreters from a postgraduate professional interpreting programme, after their eye-tracked interpreting task without live transcription in the first half and with live transcription in the second half, which cover three themes: challenges faced in interpreting with live transcription, participants' assessment of live transcription's utility during their tasks, and their strategies for attention management. The interview results were analysed in the light of established models of information processing and cognitive load. The findings were also triangulated with the eye-tracking data about the participants' interpreting behaviours and attention distribution patterns.

Literature review

Multi-taking and multimodal processing in simultaneous interpreting. Multitasking during the act of simultaneous interpreting stands as a central concern both for practitioners in the field and for academic researchers. The intricate nature of the task requires interpreters to engage in several mental operations, often juggling multiple tasks concurrently (Lederer 1981; Su and Li 2020). Probing deeper into the primary operational process of simultaneous interpreting, it can be delineated into an intricate interplay of continuous, successive, and concurrent tasks. These include, according to Lederer (1981), attentive listening, in-depth language comprehension, the intricate process of conceptualization, and drawing from cognitive memory to express the interpreted message. To elucidate further, interpreters commence by intently listening to the discourse presented by the speaker. They then embark on the challenge of comprehending this discourse, which is often in a language different from their mother tongue. This comprehension phase is not an isolated act; rather, it requires the integration of the linguistic input with their preexisting knowledge, leading to the formation of a conceptual framework in their cognitive memory. Finally, interpreters convey the message, translating their cognitive memory's stored information into articulate expression.

In simultaneous interpreting, Gile's Effort Models have been instrumental in understanding the multifaceted nature of this cognitive process (Gile 1985, 1997). Gile proposes that simultaneous interpreting can be distilled into three core components: listening and analysis (L), production (P), memory (M), plus crucial coordinating effort (C). Mathematically represented, the formula becomes SI = L + P + M + C (Gile 1985). This approach underscores the inherent mental gymnastics involved, suggesting that there is a capped mental processing capacity or energy available to an interpreter. This ceiling means that the combined exertion across listening, production, and memory must never outstrip the interpreter's total processing capability. Gile underscores the precarious equilibrium that interpreters must uphold. He posits the concept of "spill over" to describe instances where the combined cognitive demands overreach the interpreter's processing limits. In line with this, Gile postulated the "tightrope hypothesis" (Gile 1999), proposing that interpreters, especially in simultaneous settings, consistently operate at the brink of their cognitive limits. Consequently, an influx of intricate inputs, like specialized terms, proper nouns, or numbers, can strain the interpreter's capacity. This overload can instigate additional cognitive planning efforts and may, under certain conditions, cascade into what Gile terms "failure sequences" (Gile 1997, 2009). To navigate these cognitive bottlenecks and potential pitfalls, Gile suggests specific coping strategies. These include preventive tactics like notetaking, particularly when confronted with complex data or terms that may be challenging to recall or immediately translate due to linguistic constraints. Another proposed strategy involves adjusting the Ear-Voice Span (EVS). By reducing this lag, interpreters can alleviate some of the demands on their short-term memory, making the process more fluid and manageable (Gile 2009).

Seeber (2007) adapted Wickens's, (2020, 2002) Cognitive Load Model, which is three-dimensional to two-dimensional footprints for shadowing, sight translation and simultaneous interpreting. According to Wickens, (2020, 2002), the multidimensional model proposes that distinct categories and dimensions play a role in determining the efficacy of multitasking performances. The initial dimension, termed processing stages, differentiates between a communal reservoir for both perception and cognition and another distinct pool dedicated to response-related activities. The subsequent dimension, labelled processing codes, segregates tasks into manual/spatial and vocal/verbal categories. The third dimension, termed processing modalities, denotes whether the interactions occur audibly or visually. The last dimension, related to visual processing, contrasts focal from ambient visual experiences and is inherently linked to the visual modality, encompassing both spatial and verbal components. These two visual pathways are believed to access distinct resources, facilitating optimal multitasking.

Simultaneous interpreting, as described within Seeber's (2007) cognitive resource model, is an intricate multitasking activity that commands simultaneous activation of specific cognitive faculties. The model details an initial task that calls upon auditory verbal resources during both the perception and cognition stages. This is the stage where the interpreter actively listens to the spoken content and processes its meaning. Concurrently, a secondary task is initiated which harnesses vocal verbal resources, predominantly in the response phase. This enables the interpreter to vocalize a translation of the original content in real-time. One of the distinctive attributes of simultaneous interpreting, as highlighted in Seerber's model, is the overlap of auditory verbal resources during the cognition phase. This arises because the two tasks, though distinct, cooccur. As interpreters render an output, they concurrently process incoming speech. Consequently, there's a confluence at this juncture, leading to what can be visualized as a communal resource pool. This overlap is significant, as evidenced by an interference score peaking at 9, indicating a high degree of cognitive overlap and potential strain.

Seeber's conflict matrix provides a graphical illustration that quantifies the level of conflict between simultaneous tasks by aggregating the demand vectors of each sub-task and their associated individual conflict coefficients. A maximum interference coefficient of 1 signifies two tasks that are mutually exclusive in terms of resource allocation. Demand vectors are integers, including \emptyset to indicate no resource demand. When all individual demand vectors are set to 1, the point at which they intersect discloses the conflict coefficient. A baseline conflict coefficient of 0.2 represents the default competition for general capacity when tasks are time-shared. Each added dimension to the task augments this coefficient by 0.2. The overall interference score is computed by summing the demand vectors and conflict coefficients. In the specific case of simultaneous interpretation (SI), this cumulative interference score stands at 9.

In expanding the purview of cognitive studies, scholars have intensified their endeavours to explore the realm of information processing, particularly focusing on the intricacies of content processing during interpreting, seminal work on the model of multimodal information processing offers a profound insight into this. Wang (2023) posits that during interpreting, information is processed multimodally, evolving into what he terms a "hyperdiscourse" (Wang 2023, p.9). At the heart of Wang's argument is the idea that interpreters render not only the spoken words. They adeptly synthesize the speaker's array of communicative signalsfrom the explicit verbal utterances to the more subtle paraverbal nuances, and even the silent nonverbal cues. Furthermore, the interpreter is acutely aware of, and integrates, the immediate situational backdrop and the broader sociocultural intricacies inherent to the source language when engaged in simultaneous interpreting. When the multimodal information is interpreted, the resulting content is rendered in the target language, infused with a blend of verbal clarity and the resonance of paraverbal attributes. The end recipient, or listener, thus perceives this as a multi-layered hyper-discourse. This intricate output constitutes not only the interpreter's verbal and paraverbal renditions but also retains the vestiges of the speaker's inherent paraverbal and nonverbal gestures, crafting a holistic communicative experience.

Research on CAI tools and interpreters' perceptions. As the integration of information and communication technologies into the field intensifies, Fantinuoli (2017) categorizes the technological advancements into two dimensions. "Settingoriented" technologies as the first, exemplified by platforms akin to Remote Simultaneous Interpreting (RSI) systems, reshape the physical and ambient context in which interpretation unfolds. On the other hand, the "process-oriented" technologies, epitomized by instruments like Computer-Assisted Interpreting (CAI) solutions, target the intricate processes intertwined with interpreting. Process-oriented technologies in the context of interpreting refer to ICT tools and software designed to support interpreters throughout the various sub-processes of interpreting tasks (Fantinuoli 2017). Setting-oriented technologies encompass ICT tools and software that influence the external conditions under which interpreting tasks are performed or learned (Fantinuoli 2017).

Within this spectrum, platforms like Zoom, primarily recognized for their "setting-oriented" attributes, have blurred these categories. Although rooted as an online meeting tool, Zoom has metamorphosed to also embody "process-oriented" traits, such as live transcription and simultaneous interpreting functionalities. Its dual nature, coupled with its global acceptance and costefficiency relative to dedicated CAI tools, positions Zoom as an indispensable asset for modern interpreters. This synergy makes Zoom both an environmental facilitator and a process-enhancer. The recent surge in interest around CAI and RSI tools from both practitioners and academic researchers has fuelled explorations into their efficacy and impacts. A focal area of such studies has been the influence of technology on terminological and numerical interpretation accuracy. For instance, studies suggest that automatic speech recognition substantially improves the accuracy of interpreting number-dense speeches, marking improvements of over 30% (Defrancq and Fantinuoli 2021; Desmet et al. 2018; Pisani and Fantinuoli 2021).

Research probing into CAI tools has often leveraged post-task questionnaires to assess user experiences (Defrancg and Fantinuoli 2021; Desmet et al. 2018; Gacek 2015; Pisani and Fantinuoli 2021; Prandi 2015). An illustrative study by Biagini (2015) evaluated the merits of an electronic glossary, InterpretBank, vis-a-vis traditional printed glossaries. To conduct the study, two student groups underwent training: one group with InterpretBank and the other with printed glossaries. This training included a theoretical session followed by practical exercises. Post-training evaluations showed a clear preference among students for InterpretBank, citing its more efficient and smoother terminology searches. Empirical evaluations confirmed this preference, as InterpretBank users exhibited increased translation accuracy, fewer omissions, and fewer errors in translations. Moreover, the success rate of terminology searches was higher with InterpretBank than with printed glossaries. Given the notable improvements observed after limited training, it's inferred that prolonged training might optimize user performance even further. The study underscores the considerable advantages of CAI tools, specifically InterpretBank, in refining the SI process and increasing user satisfaction.

Prandi's series of studies (2015, 2017) encapsulated the same layer of the discourse, focusing on student interactions with CAI tools in educational settings. Prandi (2015) conducted a pilot study to scrutinize how interpreting students utilized CAI tools, specifically when looking up terminology during their tasks in the booth. This venture was initiated not just to observe the behavioural dynamics but also to discern how seamlessly these tools could be assimilated into the interpreting curriculum. In doing so, Prandi hoped to unearth potential hitches and, more importantly, to propose actionable solutions. The experimental framework constituted 12 MA interpreting students who were given the task of simultaneously interpreting texts heavily laden with specialized terminologies. A standout observation was the pivotal role that experience played in enabling students to seamlessly merge the tool into their booth workflows. However, this integration wasn't uniform across the board. While a segment of the students showed a propensity to lean heavily on the software, almost to the point of over-reliance, another subset found it distracting, claiming it muddied their focus during the actual interpretation process.

Later, focus shifted to assess the efficacy of the stimuli used in the data collection process. The overarching question was whether the deployment of CAI tools either led to cognitive saturation or acted as a mitigating agent, reducing cognitive load during the dual tasks of terminology search and subsequent delivery of the target text. The transcriptions from the test subjects' performances have effectively validated the suitability of the stimuli designed for the experiment. A salient outcome from this study was the evident precision offered by InterpretBank. Moreover, the glossary query emerged as the most favoured strategy among interpreters, especially when faced with the daunting task of navigating through specialized terminologies using InterpretBank.

Venturing further, Defrancq and Fantinuoli, (2021) charted the territory of automatic speech recognition (ASR) via the InterpretBank model. Their dual-method approach utilized both an

error matrix for objective accuracy measures and user questionnaires to grasp subjective experiences. Key outcomes indicated that when the ASR furnished a comprehensive transcription, especially emphasizing numerical data, there was a notable improvement in interpreting performance across almost all numerical categories. However, user engagement with ASR displayed variability, as participants only tapped into the ASR system for slightly more than half of the presented stimuli. This pattern suggests a use, wherein even though the tool proved beneficial, users exercised discretion in their engagement. Moreover, the mere presence of ASR appeared to impart a heightened sense of confidence among users. Yet, this also hinted at a possible over-reliance on the technology. This dynamic accentuates the essential equilibrium that interpreters must maintain between technological support and their inherent skills and discernment.

By analysing the interpreting performance triangulated with the eye-tracking data, Yuan and Wang (2023) investigated the cognitive processing patterns of live captioning during simultaneous interpreting on Zoom, with a specific focus on numbers and proper names, which are known to pose significant challenges. Eye movement data revealed that trainee interpreters actively searched for numbers and proper names within the live captions displayed on the screen. Their attention peaked in areas with a high concentration of these elements. Despite this focused attention, the presence of numbers and proper names did not significantly reduce attention shifts between the speaker's face and the live captioning area. According to Yuan and Wang (2023), while trainees actively look for these elements, they may shift their attention away when encountering incongruent information to manage their cognitive load. Additionally, trainee interpreters exerted more cognitive effort on segments containing numbers and proper names than on other types of information. Performance data indicated that live captioning on Zoom positively influences the accuracy of interpreting numbers and proper names, as evidenced by lower error rates. In Yuan and Wang's, (2024) study, it is found that a statistically significant difference, with a t-value of 2.479 and a p value of 0.021 (p < 0.05), indicated that enhanced performance when transcripts were available. Participants' scores increased from 58.95 without transcripts to 63.33 with live transcripts. This evidence underscores the potential of live transcription technology to bolster accuracy and reliability in interpreters' performances in RSI settings.

Summary. The study of cognitive and information processing models in simultaneous interpreting highlights the complex multitasking nature of the task, requiring interpreters to balance listening, comprehension, conceptualization, and memory retrieval within limited cognitive capacity (Lederer 1981; Gile 1985, 1997; Su and Li 2020). Gile's models emphasize the risk of cognitive overload, particularly with specialized terms and numbers (Gile 1999, 2009), while Seeber (2007) and Wang (2023) introduce the concepts of cognitive strain and multimodal processing, respectively.

Technological advancements, particularly CAI tools like InterpretBank and ASR, have significantly improved interpreting accuracy and user satisfaction (Fantinuoli 2017; Defrancq & Fantinuoli 2021; Biagini 2015; Prandi 2015, 2017). However, while many studies have assessed the effectiveness of these tools, Prandi (Alicja and Okoniewska, 2022) highlights a gap in understanding how CAI tools influence the actual interpreting process itself. Yuan and Wang (2023) addressed part of this gap through an eye-tracking experiment and analysis and found that interpreters actively search for numbers and proper names though this does not significantly reduce attention shifts. Building on this, our study aims to further address the gap with an analysis about interpreters' user perceptions about the use of ASR-generated live transcription in simultaneous interpreting. In this study, the independent variable is the utilization of live transcription, while the dependent variables include participants' responses, which vary based on different conditions and experiences. By conducting structured interviews, we investigate interpreters' user perceptions about the cognitive complexities and attention management strategies involved, using Seeber's (2007) model to calculate the specific interference score, enabling us to compare the degree of interference with other settings. This research triangulates the results by incorporating users' perceptions, with a view to providing insights into the challenges and utility of live transcription in simultaneous interpreting.

Research design

Preparation for the interpreting experiment. At the onset of the experiment, participants were introduced to the simultaneous interpreting features available on Zoom. Each function of the platform was systematically demonstrated to ensure that participants had sufficient time to understand and operate the various features. Additionally, participants had the opportunity to engage in a test session of interpreting using Zoom to further familiarize themselves with the platform's capabilities.

Participants. The study sample consisted of 13 graduating trainees from a postgraduate professional interpreting programme. These trainees, all majoring in interpreting, are prospective professionals who have undertaken two semesters of training in consecutive and simultaneous interpreting, providing a fundamental level of proficiency and competence in the discipline. The cohort included nine females and two males, all aged between twenty-two and twenty-four. Their first language (L1) is Mandarin Chinese, which is the target language for our experiment, while their second language (L2) is English. Being registered for the same course, all participants had amassed comparable interpreting practice within a classroom environment. Before the commencement of the experiment, each student had completed a minimum of six months' specialized training in simultaneous interpreting.

Speeches. The two speeches selected for the experiment—one for the task of simultaneous interpreting with live transcription and the other without a transcript—were meticulously chosen to exhibit comparable levels of complexity. Both speeches were derived from the same article on a current topic in The Economist, ensuring thematic consistency while differing in content. The first speech had a duration of seven minutes and 29 s, whereas the second speech lasted eight minutes and 39 s. These speeches were pre-recorded on Zoom by a male native English speaker at a normal delivery rate. The delivery rate for the first speech was 113 words per minute (wpm), and for the second speech, it was 112 wpm.

In the first speech, which had a total word count of 855, numbers constituted approximately 2.1% of the total word count (18 occurrences), topic-related terms accounted for about 8.1% (69 occurrences), and proper nouns comprised around 2.0% (17 occurrences). In the second speech, with a total word count of 913, numbers represented approximately 2.0% of the total word count (22 occurrences), topic-related terms constituted about 5.3% (49 occurrences), and proper nouns made up approximately 1.6% (15 occurrences).

We also quantified the complexity of the speeches using various metrics. The Flesch Reading Ease score, a commonly used

readability index, was calculated for each speech. Additionally, we also computed the lexical density, a ratio of the number of unique words to the total word count, as an alternate measure of complexity. The data revealed a Flesch Reading Ease score of 52.9 for the first speech and 52.75 for the second. Concurrently, the lexical density for the first speech stood at 0.58, and for the second, it was 0.54.

These findings indicate that the two speeches exhibit nearly identical levels of length, delivery rate, and complexity, particularly in terms of numbers and proper nouns. This uniformity effectively eliminates speech complexity as a confounding factor, thereby enhancing the generalizability of the results. By ensuring that the two speeches share comparable levels of complexity, we have ensured that the study's outcomes are primarily influenced by the presence or absence of live transcription during simultaneous interpretation, rather than by any differences in speech complexity.

An eye-tracked interpreting experiment. Participants were oriented to the actual experiment next, which included the task of simultaneous interpreting on Zoom and an interview that explored multimodal processing in simultaneous interpreting within the same platform.

Although the participants had not received explicit training on the usage of live transcription across any platform, they had previous exposure to on-screen information. They were briefed that the interpreting task would require simultaneous interpreting with and without live transcription. The entire study included the English-Chinese interpreting tasks, the collection of eye movement data, and post-experiment interviews. As Yuan and Wang (2023) presented the eye-tracked interpreting experiment and analysis about the interpreting performance, this study focuses on interpreters' user' perceptions and its triangulation with the eyetracking processes and interpreting performance data. Section 3 explicitly explains the experiment involving the interviews and interpreting recordings.

The primary focus of our study was to maintain ecological validity by emulating real-world training conditions typically employed in professional-level programmes. The speeches for the task were prepared and delivered extemporaneously, implying that though they were planned in advance, their presentation was spontaneous. Recording the speeches ensured uniformity in delivery, preventing any influence on the results' generalizability. To maintain consistency within the sample, the participant pool was kept as homogeneous as possible regarding their experience with simultaneous interpreting. Thus, only programme students who had received an equivalent amount of training were selected for the experiment.

Post-experiment interviews. After the interpreting experiment, the participants were invited to a structured interview including 17 questions. The questions in the interview were designed by the authors to evaluate the subjects' multimodal processing during the task, which involved both the presence and absence of live transcriptions. It aims to investigate the effects of live transcription on the performance of simultaneous interpreting and how interpreters perceive them. The interview questions centred around three overarching themes:

1. Participants were asked to identify the specific segments of the speech that were difficult to interpret and the factors that contributed to the difficulty.

The first set of questions aim to gain a comprehensive understanding of the challenges that participants are faced during the interpreting task and the potential factors that have influenced their performances. 2. Participants were asked to provide their opinion on whether the live transcription was helpful to their interpreting process and how.

The second set of questions are designed to evaluate the participants' perception of the effectiveness of the live transcription feature provided by Zoom in facilitating simultaneous interpreting. In addition to the binary 'yes' or 'no' response options to indicate the usability of the live transcription, especially in relation to terminologies and numbers, a 5-point Likert scale was also employed to determine the extent to which the live transcription served as either a helpful tool or a distraction for the participants. Furthermore, participants were encouraged to elaborate on the reasons behind their preference for or against the use of live transcription during the task. The results were verified with the triangulation of the participants' performances and eye-tracking data.

3. Participants were asked to retrospect their attention division during the interpreting process.

Participants were asked to describe the proportion of attention they allocated to each mode (listening and reading transcription) during the interpreting task. They were also asked to indicate whether they were more focused on the speaker's facial expressions and, if so, whether there were any conflicts between the nonverbal cues and the live transcript. Participants were encouraged to provide explanations for their responses to gain a deeper understanding of the factors that influenced their focus during the interpreting task. The primary aim of this theme of the interview was to gain insights into how the participants managed their attentional resources and the extent to which the live transcription or facial expressions affected their attentional focus. By exploring these issues, the questionnaire aimed to provide insights into the participants' ability to process multimodal information.

Procedure for data analysis. This study collates both performance and perceptual data, analysing these through quantitative and qualitative methodologies. Performance data served to assess the actual efficacy and efficiency of live transcription, given that participants' perceptions may not reliably indicate these aspects. With respect to the error rate in interpreting numbers and terminologies, both with and without live transcript, data sets were compared using a paired t-test. This statistical tool allows for the comparison of the means of two dependent groups to determine if there is a significant difference between them. The data visualization was enhanced through the use of boxplots, offering a comprehensive summary of the data set's minimum, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile), and maximum values. To understand how the participants process multimodal information with live transcription, the study integrates the model of multimodal information processing in SI by Wang (2023), and Cognitive Resource Footprint and Conflict Matrix by Seeber (2011, 2007).

Research questions

In accordance with the overall experimental design, particularly the post-experiment interviews, the research seeks to explore trainee interpreters' perceptions regarding the use of ASRgenerated live transcription to facilitate English-Chinese simultaneous interpreting. The following research questions are formulated:

RQ1. What challenges do trainee interpreters face when using live transcription during simultaneous interpreting?

RQ2. How do trainee interpreters perceive the usefulness of live transcription in the interpreting process?

RQ3. How do interpreter trainees allocate their proportion of attention to each mode (between listening and reading transcription) during the interpreting task?

Results

Challenges in simultaneous interpreting with live transcription. Participants' experiences varied in terms of which segments in the speech were particularly challenging to interpret. While most found difficulty with specific aspects of "numbers" and "proper nouns", some others pointed to challenges with broader sections of the speech, such as "the part with live transcription" or "the first part of speech". Several participants mentioned deeper into the "other reasons" that posed challenges in interpreting. Many reported experiencing an intensive cognitive load, especially after extended periods of work. This was often compounded by a compulsion to continuously monitor the live transcriptions displayed on the screen. The task of dividing their attention between the speaker's speech and the live transcriptions further exacerbated the situation. Some participants also grappled with long sentences or were distracted by the mere presence of words from the live transcriptions. Slow speech speeds posed another hurdle, making it challenging to keep pace. There were also mentions of a felt inadequacy in active listening skills, which, combined with dense information, made quick processing a daunting task. Furthermore, lapses in concentration occasionally interrupted their ability to consistently follow the speaker. The results highlight the myriad challenges faced by interpreters during simultaneous interpreting with live transcription. The challenges span cognitive, technical, and skill-related aspects, with a particular emphasis on managing live transcriptions.

Utility of live transcription in simultaneous interpreting

Usefulness of live transcription. The data of participants' views on the efficacy of the live transcription for simultaneous interpretation (SI) was split: seven of the thirteen respondents (54%) felt SI was more challenging with live transcription, whereas six (46%) felt the opposite. Despite this division regarding difficulty, 69% of participants believed live transcription aided their interpretating tasks. In contrast, 23% found it disruptive, and one participant (8%) opted not to use it at all. On a 5-point Likert scale, those deeming the feature useful assigned it an average score of 3.11, suggesting its utility was perceived as moderate. Those finding it distracting scored it higher at 4.33, signalling a significant hindrance.

To further validate the efficacy of live transcriptions for SI, we analysed error rates in interpreting numbers and terminologies commonly challenging aspects as cited by participants. This involved comparing error rates with and without the transcripts. A visual assessment was first conducted using boxplots. Boxplots, also known as whisker plots, provide a graphical view of the spread and skewness of data, illustrating the distribution of error rates. The box in the boxplot displays the interquartile range (IQR), representing the middle 50% of the data. The line inside the box shows the median error rate. Outliers can be identified as data points outside the whiskers of the boxplot. From our boxplots, it was evident that the error rates when using live transcription were clustered towards the lower end. In contrast, without live transcripts, error rates spread towards higher values, indicating increased errors in this scenario.

To statistically verify the observed differences from the boxplots, a paired t-test was conducted. This test is appropriate for comparing two related samples, in our case, error rates with and without the use of live transcription from the same group of participants. Prior to conducting the paired t-test, the differences in error rates between interpreting tasks with and without live transcription were subjected to a normality test and an outlier detection analysis to ensure the validity of subsequent statistical inferences. The Shapiro-Wilk test was applied to assess the normality of the differences, yielding a statistic of 0.962 and a p-value of 0.819. This result suggests that the differences in error rates are normally distributed, as the p-value is well above the conventional alpha level of 0.05, thereby not rejecting the null hypothesis of normality. Furthermore, an outlier analysis using the interquartile range (IQR) method identified no outliers in the data, indicating that all differences fall within an acceptable range and that the results are not skewed by extreme values.

Subsequently, the *t* test assesses if the mean difference between these two sets of observations is statistically significant. The calculated t-statistic was -7.40, which signifies that error rates were, on average, lower when using the live transcripts. The extremely small *p* value of 1.37×10^{-5} (much less than the standard significance level of 0.05) provides compelling evidence against the null hypothesis, strongly suggesting that the observed difference is statistically significant and not due to random chance. To further bolster our findings, we compared our results with an the study by Yuan and Wang (2023). The congruence between our study and theirs adds reliability to our findings.

In essence, despite the associated challenges, visual and statistical analyses strongly support the beneficial impact of live transcripts, specifically in minimizing errors in interpreting numbers and terminologies (Fig. 1).

Utilization of live transcription. Regarding those participants who perceived live transcription as a beneficial tool, further data was gathered on how they specifically employed this feature. 40% of these participants reported that they primarily used live transcription when they had difficulty discerning the auditory message of the speech. Equally, another 40% indicated that they used live transcription as a means to interpret the entire transcript, suggesting that these participants were engaging in sight translation alongside their simultaneous interpretation of the speech. A minority, only 20%, reported using live transcription exclusively as a prompt or guide during interpretating tasks (Fig. 2).

In contrast, of the participants who found live transcription a distraction, three participants noted the value of the transcript in managing dense information and identifying key elements. Conversely, some found live transcription a hindrance, citing the discord between interpreting rhythms from the transcript versus the source speech. Others mentioned the inherent latency of live captioning, with official sources by Zoom citing a delay of 6–10 s. A few participants identified challenges with sight interpreting, either due to the latency or personal limitations.

Given these varied experiences, while many find live transcription beneficial, its impact varies by individual. To deepen our understanding, subsequent investigations will involve eye-tracking analyses, assessing engagement with live transcriptions and its correlation with performance in the next section (Fig. 3).

Attention division in simultaneous interpreting with live transcription. In simultaneous interpreting, the advent of live transcriptions introduces an additional dimension of cognitive demand. Participants in our study revealed a spectrum of attention distribution strategies when navigating between listening to a speech and concurrently reading a live transcript. While a substantial number of participants leaned heavily on auditory processing, sometimes directing up to 85% of their focus on listening, others found greater reliance on the written content, dedicating as much as 80% of their attention to the transcript. This difference emphasizes the individualized strategies interpreters develop over time and experience.



Fig. 1 Boxplots of error rates during SI tasks with and without live transcript. The boxplots in Fig. 1 compare error rates in simultaneous interpretation (SI) tasks involving numbers and terminologies, with (left boxplot) and without (right boxplot) live transcription. Each boxplot shows the interquartile range (IQR) capturing the central 50% of error rates, with the median error rate depicted by the line inside the box. Lower error rates are indicated by the clustering of data points towards the lower end of the scale with live transcription, whereas a higher spread of data points towards the upper end is observed without transcription. Outliers, should they exist, are noted by data points that lie beyond the whiskers.



Fig. 2 Users' perception on how they used live transcription if they found it useful. This pie chart displays the various ways interpreters employed live transcription, with 40% using it when auditory challenges arose, another 40% for full sight translation, and 20% as an occasional aid, reflecting diverse strategies and reliance on the tool.

The dual task of balancing auditory and visual stimuli was not without challenges. Some participants expressed the struggle in maintaining a balance, with one participant initially trying to equally divide attention, but later finding themselves veering towards the transcript more. This highlights the cognitive strain and adaptability required for such multitasking endeavours. There were also unique approaches among the cohort. For instance, one participant confessed to adopting a purely sight translation method during the initial phases, bypassing the auditory component entirely.

When probing into how participants divided their attention between the live transcript and the speaker's facial expressions, we asked four specific questions: (1) Did you watch the speaker's facial expression when interpreting? (2) Was there a conflict in processing this and the live transcription? (3) Did you pay more attention to the speaker's facial expression or the live transcription? (4) Why did you choose to do so?

Our findings from these questions showed intriguing patterns. The observation of facial expressions shows that out of 13 participants, a significant majority of 8 actively monitored the speaker's facial expressions. This behaviour underscores the invaluable role of non-verbal cues in the interpretative process, suggesting that facial expressions enhance the richness of verbal content, providing an added dimension of context not always evident in the words alone. However, the multitasking nature of simultaneous interpreting became evident when considering potential conflicts between live transcription and facial cues. While 10 or the majority of participants managed to coordinate both inputs, three were faced with cognitive challenges. This discrepancy sheds light on the varying adaptability among interpreters; while many have refined their ability to process concurrent stimuli, some find the balancing act challenging. In terms of attentional priorities, it is observed that 11 participants favoured the live transcription over the speaker's facial expressions. In contrast, a minority, only 2 participants, placed more weight on facial expressions, hinting at a potentially more holistic or visually inclined interpretation style.

Unpacking the reasons behind these preferences based on the participants' response was enlightening. For many, past experiences and entrenched habits played essential roles. While some had cultivated a preference for auditory stimuli, others had become more attuned to written content over the years. Additionally, the dynamism and precision of the live transcript emerged as a recurring theme. Its evolving nature, especially with the inclusion of crucial terms or numbers, provided a sense of reliability. However, the character of the speaker in the video also influenced decisions: less expressive speakers naturally shifted the interpreters' focus towards the richer, more informative live transcript.

Discussion

Heightened cognitive load in multimodal information processing in simultaneous interpreting with live transcription. Responding to the first research question on the challenges encountered during simultaneous interpreting with live transcription, interpreters face an intricate web of linguistic, cognitive, and technological obstacles. Specific issues arise from language segments, cognitive overload from multitasking, distractions due



Fig. 3 Participants response to attention division between speaker's facial expression and live transcription. The bar chart in Fig. 3 depicts interpreters' attention allocation between observing the speaker's facial expressions and consulting live transcription during interpretation tasks. A majority of 8 out of 13 participants watched the speaker's facial expressions, while 10 managed to integrate both facial cues and transcription without conflict. Conversely, 3 participants struggled with this multitasking. The chart further shows that 11 participants predominantly focused on live transcription, with only 2 favouring facial expressions.

to live transcriptions, and the arduous task of maintaining focus throughout. The management of live transcriptions, with its dual nature of being a potential distraction and aid, emerges as a significant concern.

Utilizing Wang's (2023) model of multimodal information processing as an analytical lens, it is confirmed that interpreters process the information multimodally as a "hyper-discourse" (Wang 2023, p.9). In simultaneous interpreting with live transcription, participants listened to the speech, absorbing verbal information that contained the speech's message content. Simultaneously, they received paraverbal information, such as pitch, tone, speech speed, pauses, and hesitations. Zoom also displayed the speaker's image, conveying nonverbal information such as facial expressions, gestures, and postures. When live transcription is activated, another layer of information-the visual representation of the transcribed speech messages-is presented to the interpreters for processing. This could potentially augment the cognitive load, as additional information needs processing for output, but it could also serve as a useful tool, as discussed in the following section. With all these multimodal information inputs, participants produced verbal and paraverbal information for the listener, though without nonverbal information, given their online virtual booth setup.

Live transcription as an aid to simultaneous interpreting. In response to the second research question, it has been observed that the integration of real-time transcription within the context of simultaneous interpretation, has introduced a novel dimension to this practice. Given the complexities inherent in interpreting, especially within language pairs such as English/Chinese, the inclusion of an additional supportive tool can be helpful. The empirical data indicates that participants strategically employ live transcription, leveraging its functionalities to effectively address a diverse array of challenges encountered during interpreting tasks. Live transcription emerges not just as an ancillary tool, but as an instrument that can be woven into the very fabric of the interpreting process.

Our inquiry explored the intricacies of cognitive processes, revolving around the unique way of information processing that interpreters follow during the practice of simultaneous interpreting, which is further enhanced by the integration of live transcription. Drawing insights from scholars like Yang (2010), Dawrant (1996), Wang and Gu (2016), Alexieva (1983) and (Li 2014), interpreters engage in a systematic information processing approach, adhering to the linear progression of the source speech. This entails the interpretation of smaller segments of the source text (ST) in their original order, followed by the reproduction of subsequent segments, employing a variety of techniques to accomplish this task (Yang 2010). The distinct attributes of the English/Chinese language pair introduce an additional layer of complexities for interpreters (Wang and Gu 2016). Given these factors, the interpretation of propositional information emerges as a notable challenge. Alexieva (1983) emphasized that successful simultaneous interpretation hinges on the interpreter's capability to identify propositional semantic constructs embedded within the text they are actively handling. She asserted that lacking this skill, interpreters would be limited to generating fragmented interpretations primarily composed of isolated words.

According to the users' perception, the convergence of cognitive process with live transcription encompasses visual support in source speech assimilation, anchoring semantic segments through transcription, and facilitating accurate targetlanguage formulation. In more precise terms, as interpreters engage in the initial assimilation of the source speech, the presence of live transcription enhances this auditory encounter through a visual representation. The existence of transcribed text aids in marking key segments or 'signposts' within the speech. This amalgamation of auditory and visual inputs assists interpreters in erecting a mental framework of the discourse, thereby possibly enhancing the fluidity of the interpretation process. Subsequently, when grappling with the intricate task of discerning semantic constructs or even propositional information, live transcription assumes the role of a reference point. It provides a tangible point of reference, allowing interpreters to validate their comprehension. The ultimate act of formulating interpretations is fortified by live transcription, which stands as a resource.

Attentional dynamics and cognitive allocation in simultaneous interpreting with live transcription. The analysis on attention division amidst simultaneous interpretating using live transcription



Fig. 4 Cognitive resource footprint for SI with live transcription (Adapted from Seeber, 2007, 2017). The figure illustrates the cognitive resource allocation in simultaneous interpreting with live transcription, highlighting the various cognitive demands. Visual-spatial resources are engaged when interpreters visually locate transcription for auditory comprehension. Attention is then split between visual (reading) and auditory (listening) inputs, involving visual-verbal resources. During interpreting, perceptual-cognitive processes activate auditory-verbal, visual-verbal, and cognitive-verbal resources, while vocal-verbal resources are used in the production.

extends our understanding of cognitive dynamics, fostering a richer understanding of the users' experience.

In dissecting the third research question about attention division, what emerges as striking is the heightened focus on live transcription during instances marked by the density of numbers and proper nouns. When analysed in conjunction with selfreported interview responses, the findings drawn from eyetracking data (Yuan & Wang, 2023, 2024) corroborated the participants' alignment towards live transcription. This magnified attention, explicitly underlined by participants in qualitative interviews, mirrors the inherent challenges these segments pose and the resultant cognitive emphasis on ensuring precise interpretation. On the flip side, when the speech is dense, the scant attention on the speaker's facial expressions speaks to the complexities of assimilating both live transcription and nonverbal cues. Here, the cognitive bandwidth is majorly reserved for decoding the textual elements, with facial cues being momentarily side-lined. Yet, the elasticity of the interpretation process is illuminated when the live transcription isn't densely packed. Participants' pivoting focus to the speaker during these phases underscores the continual juggling of attention, underscoring the interpreter's adaptability.

To gain insights into the cognitive complexities experienced by interpreters, especially when they incorporate a new tool such as live transcription, we apply Seeber's cognitive resource footprint and conflict matrix to discern the types of cognitive resources engaged and the extent of their interference with one another.

From our analysis, alongside simultaneous interpreting operations, when interpreters work with live transcription, we can categorize participant responses from interviews into two distinct groups. Some interpreters engage in sight translation, allocating a significant portion of their attention to live transcription. The majority of participants employ live transcription as a supplementary aid during challenging moments of auditory comprehension, particularly when confronted with numbers, proper nouns, and specialized terminologies. The eye-tracking results (Yuan and Wang 2023) validate this behaviour, offering a comprehensive portrayal of the participants' dynamic attentional shifts during periods of interest (POIs) characterized by an abundance of numbers and proper names. Interestingly, all but one participant confirmed that their attention oscillates among live transcription, observing the speaker's facial expression, and active listening.

It's important to recognize that the footprint outlined in Fig. 4, adapted from Seeber (2007, 2017), specifically represents instances where interpreters are simultaneously utilizing live transcription and attending to the speech. As illustrated in Fig. 4, during simultaneous interpreting with live transcription, two tasks converge. The first task of listening, reading, and comprehension activates auditory-verbal, visual-verbal, and

			listening & comprehension								
				perceptual				cognitive		response	
			vector	ø	Ø	Ø	1	ø	1	Ø	Ø
production & monitoring		demand		visual spatial	visual verbal	auditory spatial	auditory verbal	cognitive spatial	cognitive verbal	response spatial	response verbal
	perceptual	1	visual spatial	0.8	0.6	0.6	0.4	0.7	0.5	0.4	0.2
		1	visual verbal	0.6	0.8	0.4	0. 6	0.5	0.7	0.2	0.4
		Ø	auditory spatial	0.6	0.4	0.8	04	0.7	0.5	0.4	0.2
		1	auditory verbal	0.4	0.6	0.4	0.8	0.5	0.7	0.2	0.4
	cognitive	Ø	cognitive spatial	0.7	0.5	0.7	0.5	0.8	0.6	0.6	0.4
		1	cognitive verbal	0.5	0.7	0.5	0.7	0.6	0.8	0.4	0.6
	response	Ø	response spatial	0.4	0.2	0.4	0.2	0.6	0.4	0.8	0.6
		1	response verbal	0.2	0.4	0.2	0.4	0.4	0.6	0.6	1.0
Total interference score = demand vectors + conflict coefficients. = $(1+1+1+1+1+1+1)+(0.8+0.7+0.4+0.7+0.8+0.6+0.4+0.6+0.5+0.7)$.											

cognitive-verbal resources during the perceptual-cognitive stage. The second task, centred around the production and monitoring of translations, draws upon vocal-verbal, auditory-verbal, cognitive-verbal, visual-spatial (highlighted in blue), and visual-verbal resources across the response and perceptual-cognitive stages.

The conflict matrix (Table 1) specifically quantifies interference in the demanding context of simultaneous interpreting that incorporates live transcription. Demand vectors, indicated by integers ("Ø" for no demand and "1" for demand), measure resource utilization. When interpreters concurrently listen to the speech, read the live transcription, and engage in interpretation, the complexity of resource utilization intensifies. In this scenario, listening and comprehension tasks require a total demand vector of 2, reflecting the use of auditory-verbal and cognitive-verbal resources essential for processing the speech. Production and monitoring tasks, particularly with live transcription, have a demand vector of 5, involving multiple resource types as interpreters manage transcription and screen-based information.

Conflict coefficients quantitatively measure the level of interference between two tasks that utilize the same or overlapping cognitive resources. These coefficients, integral to calculating the total interference score, range from a low of 0.2 to a high of 1.0. In Table 1, where the interaction of resource demands is highlighted in dark grey, the conflict coefficients for various task interactions—such as those involving auditory-verbal, cognitive-verbal, and visual-verbal resources—are recorded as 0.4, 0.5, 0.6, 0.7, 0.8, 0.7, 0.7, 0.8, 0.4, and 0.6. These values reflect varying degrees of interference, with higher values indicating greater competition for shared cognitive resources.

The sum of these conflict coefficients, combined with the total demand vectors for the tasks involved in interpreting while managing live transcription, results in an overall interference score of 13.2. This score represents a significant increase from the baseline score of 9 observed in standard simultaneous interpreting without transcription aids. The marked rise in the interference score highlights the additional cognitive burden placed on interpreters who must juggle listening, reading, and translating in real time—a complex cognitive endeavour that demands careful management of overlapping resource demands.

Based on the response provided by the participants describing how they use live transcription during the interpreting, the cognitive load is a dynamic parameter, fluctuating throughout the interpreting process based on the cognitive resources in play. We postulate that the cognitive load peaks with the total interference score when interpreters are tasked with reading transcription, listening, and interpreting concurrently. This surge could, in some instances, escalate to cognitive overload. However, once the transcription information is successfully acquired, the cognitive load is anticipated to revert to its baseline.

Further supporting this observation, Participants' response from interviews further substantiates this dynamic cognitive load fluctuation during the interpreting process. For those with difficulties hearing, live transcription becomes a crucial tool, allowing them to rely more on visual-spatial resources to comprehend the message. In such cases, the interference score typically ranges between 8.8 and 9. However, when participants use live transcription to intermittently verify their understanding, check specific terms, or ensure interpretation accuracy, the interference spikes to 13.2 due to the simultaneous recruitment of auditory-verbal, visual-verbal, and cognitive-verbal resources during the perceptual-cognitive stage, and additionally auditoryverbal, cognitive-verbal, visual-spatial (highlighted in blue), and visual-verbal resources during the response and perceptualcognitive stages.

Once the necessary information, such as terminologies and numbers, is captured, the cognitive load is alleviated, reducing the interference score to 8.8. This reduction occurs as the cognitive effort required to process auditory stimuli is augmented by visual cues, which help streamline the cognitive process. As interpreters switch between visual and auditory inputs, their cognitive focus divides, thereby moderating the interference level to a score typical of standard simultaneous interpreting, which is 9. This highlights the adaptive nature of cognitive resource allocation in response to varying demands during interpreting.

For a subset of participants who choose to interpret entirely from the transcript rather than listening to the speech, the interference score remains consistently at 8.8. This score indicates a reduced cognitive burden as the primary reliance on visual input reduces the need for auditory processing, underscoring the varied impact of live transcription on cognitive load during interpreting tasks.

Conclusion

This study explored trainee interpreters' perception about the utility of ASR-generated live transcription in facilitating English-Chinese simultaneous interpreting. Through structured interviews, we assessed interpreters' multimodal processing and attention allocation during tasks, particularly in segments involving numbers and proper nouns. The findings reveal mixed perceptions: 69% of participants found live transcription beneficial, especially when auditory comprehension was compromised, attributing a moderate utility score of 3.11 on a 5-point Likert scale. Conversely, 23% considered it disruptive, assigning it a hindrance score of 4.33, while 8% abstained from using it altogether. Notably, 40% of interpreters used it primarily for difficult auditory segments, and another 40% employed it for complete sight translation of transcripts. Only a minority (20%) used it strictly as a supplementary prompt. The cognitive load, as quantified with Seeber's conflict matrix, markedly increases with the use of live transcription, with total interference scores spiking from 9 to 13.2. This underscores the tool's varying impact: while it serves as a primary resource in challenging auditory contexts and sight translation, it also acts as a strategic aid in less demanding scenarios. Despite the challenges, live transcription's role in enhancing interpretive accuracy highlights its potential as a valuable asset in the evolving landscape of digital communication tools in simultaneous interpreting.

The findings suggest that interpreter training programmes should include modules on effectively managing multimodal inputs and cognitive resources, while technology developers might focus on enhancing transcription accuracy and introducing customizable features for real-time use by interpreters. Professional associations can also use these insights to develop guidelines for using technological aids in interpreting. Additionally, the study highlights the need for further research to explore live transcription's impact across various settings and language pairs, and to analyse cognitive processes more deeply, potentially using eye-tracking to understand attention distribution and manage cognitive overload. Understanding these aspects can help improve interpreter well-being by developing practices that mitigate fatigue and cognitive overload, thereby promoting better mental health and job satisfaction.

This study also has several limitations. As the study's findings are drawn primarily from graduating trainee interpreters, the findings based on the experiment on and interviews with them may not be the same as those with seasoned professionals. While they offer insights into the emerging generation's adaptation to technology, it's essential for future studies to involve experienced professional interpreters for a more comprehensive picture. Meanwhile, the participants' prior experience and comfort level with digital tools like ASR and live transcription could influence their cognitive load and adaptation strategies. Those unfamiliar might face more challenges than those accustomed to such platforms. Additionally, the study focuses on Zoom's live transcription tool. The findings might not be directly applicable to other platforms or future versions of Zoom with enhanced features.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 21 November 2023; Accepted: 29 January 2025; Published online: 08 February 2025

References

- Alexieva B (1983) Compression as a means of realisation of the communicative act in simultaneous interpreting. Fremdsprachen 27(4):233–238
- Alicja SRFM F, Okoniewska M (n.d.). A Paper on the Conference Panel" Inbooth CAI Tool Support in Conference Interpreter Training and Education. (2022)
- Biagini G (2015) Glossario cartaceo e glossario elettronico durante l'interpretazione simultanea: Uno studio comparativo. Università Di Trieste MA Thesis
- Dawrant A (1996) Word order in Chinese-English simultaneous interpretation: An initial exploration. Unpublished MA Thesis, Fu Jen University
- 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 33(1):73–102
- Desmet B, Vandierendonck M, Defrancq B (2018) Simultaneous interpretation of numbers and the impact of technological support. In *Interpreting and technology* (pp. 13–27). Language Science Press
- Fantinuoli C (2017) Computer-assisted interpreting: Challenges and future perspectives. Trends in E-Tools and Resources for Translators and Interpreters, 153–174
- Fantinuoli C (2018) Interpreting and technology: The upcoming technological turn. Interpreting and Technology, 1–12
- Gacek M (2015) Softwarelösungen für DolmetscherInnen. Vienna: Uniwien. [Pobrane z Http://Othes. Univie. Ac.at/35667/1/2015-01-16_0607925.Pdf. Data Ostatniego Dostępu: 01.09. 2020]
- Gerver D (1969) The effects of source language presentation rate on the performance of simultaneous conference interpreters. Interpreting Stud Read 2002:52–66
- Gile D (1985) Le modèle d'efforts et l'équilibre d'interprétation en interprétation simultanée. Meta: J Des Trad 30:44. https://doi.org/10.7202/002893ar
- Gile D (1997) Conference interpreting as a cognitive management problem. Appl lied Psychol-Lond-Sage 3:196–214
- Gile D (1999) Testing the Effort Models' tightrope hypothesis in simultaneous interpreting-A contribution. HERMES-J Lang Commun Bus 23:153-172
- Gile D (2009) Basic concepts and models for interpreter and translator training 8. John Benjamins Publishing
- Kirchhoff H (1976) Simultaneous interpreting: Interdependence of variables in the interpreting process, interpreting models and interpreting strategies. The Interpreting Studies Reader, 111–119
- Lederer M (1981) La traduction simultanée: Expérience et théorie. *Paris: Minard Lettres Modernes*
- Li X (2014) Sight translation as a topic in interpreting research: Progress, problems, and prospects. Across Lang Cult 15(1):67–89
- Pisani E., Fantinuoli C (2021) Measuring the impact of automatic speech recognition on number rendition in simultaneous interpreting. *Empirical Studies of Translation and Interpreting: The Post-Structuralist Approach*, 181–197
- Prandi B (2015) Use of CAI tools in interpreters' training: A pilot study. Proceedings of Translating and the Computer 37
- Prandi B (2017) Designing a multimethod study on the use of CAI tools during simultaneous interpreting. Translating Computer 39:76–113
- Seeber KG (2007) Thinking outside the cube: Modeling language processing tasks in a multiple resource paradigm. *Eighth Annual Conference of the International Speech Communication Association*
- Seeber KG (2011) Cognitive load in simultaneous interpreting. Existing theories— N. models Interpreting 13(2):176–204
- Seeber KG (2017) Multimodal processing in simultaneous interpreting. In *The Handbook of Translation and Cognition* (pp. 461–475). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119241485.ch25
- Su W, Li D (2020) Exploring processing patterns of Chinese-English sight translation: An eye-tracking study. Babel 66(6):999–1024
- Wang B (2023) Exploring information processing as a new research orientation beyond cognitive operations and their management in interpreting studies: Taking stock and looking forward. Perspectives 31(6):996–1013
- Wang B, Gu Y (2016) An evidence-based exploration into the effect of languagepair specificity in English-Chinese simultaneous interpreting. Asia Pac Translation Intercultural Stud 3(2):146–160
- Wickens CD (2002) Multiple resources and performance prediction. Theor Issues Ergonomics Sci 3(2):159–177
- Wickens CD (2020) Processing resources and attention. In Multiple task performance (pp. 3–34). CRC Press
- Yang C (2010) A study of information processing in interpreting. Nankai University Press Tianjin
- Yuan L, Wang B (2023) Cognitive processing of the extra visual layer of live captioning in simultaneous interpreting. Triangulation of eye-tracked process

and performance data. Ampersand 11:100131. https://doi.org/10.1016/j. amper.2023.100131

Yuan L, Wang B (2024) Eye-tracking the processing of visual input of live transcripts in remote simultaneous interpreting: Preliminary findings. Forum Rev Int d'interprétation et de Trad / Int J Interpretation Transl 22(1):118–144. https://doi.org/10.1075/forum.00038.yua

Author contributions

In this co-authored paper, the first author conducted data analysis, drafting and revision of the manuscript; the corresponding author completed conceptualisation of the project and research design (including the interview questions), and revised the drafts substantially for intellectual content; the third author conducted the data collection.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval for this research was granted by the Arts, Humanities and Cultures Committee of the University of Leeds under approval number FAHC 21-092, dated 21/11/2023. All research activities, including study design and data collection methods, were conducted in accordance with the Declaration of Helsinki and comply with local regulations.

Informed consent

Written informed consent was obtained from all participants on 26/11/2023 by Lu YUAN, the third author. The consent forms ensured confidentiality and detailed the purpose of the research, the use of data for research and publication, and any associated

risks. Anonymity of participants is strictly maintained through measures designed to secure and anonymize data.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-04492-w.

Correspondence and requests for materials should be addressed to Binhua Wang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/ licenses/by/4.0/.

© The Author(s) 2025