



Integrating automatic speech recognition into Chinese–Portuguese simultaneous interpreting

Meng Guo

University of International Business and Economics
Beijing, China

gm@uibe.edu.cn ✉

<https://orcid.org/0000-0001-5772-2272> 

Lili Han

Macao Polytechnic University
Macao, China

hanlili@mpu.edu.mo ✉

<https://orcid.org/0000-0002-8995-2301> 

Marta Teixeira Anacleto

University of Coimbra
Coimbra, Portugal

marta@fl.uc.pt ✉

<https://orcid.org/0000-0002-7568-1942> 

Abstract: The rapid development of automatic speech recognition (ASR) has created new opportunities for technology-enhanced simultaneous interpreting, yet empirical research has focused largely on English-paired combinations. This study examines the integration of real-time automatic transcription in Chinese–Portuguese simultaneous interpreting, investigating its effects on interpreter performance and on interpreters' interaction with and perceptions of the technology. Ten first-year MA students completed simultaneous interpreting tasks with and without ASR-generated transcription in a mixed-methods design. Performance was assessed across four dimensions (information completeness, fluency, target language quality, holistic performance), with ear-voice span measured to capture temporal processing patterns. Qualitative data were collected through questionnaires and retrospective interviews. Results revealed that transcription support significantly improved information completeness but yielded no significant effects on fluency, target language quality, or holistic performance. Participants reported allocating, on average, 60% of their attention to the visual display, with ear-voice span nearly doubling when transcription was available. Perceptions of the tool were generally positive, though concerns were raised regarding cognitive fatigue and error propagation. Participants preferred continuous transcription with highlighted problem triggers, partly diverging from findings in other language pairs. Overall, the results show that ASR-generated transcription offers selective benefits while introducing risks of visual over-reliance, underscoring the need for targeted training and customizable, user-centered interface design in technology-supported simultaneous interpreting.

Keywords: simultaneous interpreting; automatic speech recognition; real-time transcription; Chinese–Portuguese; computer-assisted interpreting.



I. Introduction

The relationship between China and Portuguese-speaking countries has intensified considerably in recent decades, generating unprecedented demand for language mediation services across the Chinese–Portuguese pair—a language combination that bridges approximately 1.4 billion Chinese speakers and over 250 million Portuguese speakers across four continents. This sustained growth in cross-cultural contact has stimulated increasing scholarly attention to translation and interpreting practices for this language pair, giving rise to several distinguishable lines of inquiry.

One established research strand encompasses cultural and literary studies examining the circulation and reception of texts between Chinese and Lusophone contexts (Huang & Zhang, 2025; Yu et al., 2024), along with linguistic investigations addressing lexical, terminological, and contrastive dimensions of the two languages (Luo & Zhang, 2025; Sun, 2017). Two more recent orientations have emerged. The first explores cognitive aspects of language mediation. Building on translation process research, Silva et al. (2015, 2017) used eye-tracking and keystroke logging to examine cognitive effort in Chinese–Portuguese translation and post-editing, while Han et al. (2023) extended this line of inquiry to interpreting. The second reflects the growing intersection of technology and language services, with studies examining automated translation quality assessment using large language models (Jiang et al., 2024) and the development of machine translation corpora and systems for the Chinese–Portuguese pair (Amaro & Pires, 2024; Chao et al., 2018; Liu et al., 2018).

Despite this expanding body of research, significant gaps remain. Simultaneous interpreting (SI)—commonly defined as the near-immediate reformulation of a speaker’s message into another language while the source speech continues (Gile, 2009; Pöchhacker, 2016)—has received remarkably limited empirical attention in the Chinese–Portuguese context. This omission is striking given SI’s centrality in diplomatic, commercial, and institutional communication involving China and Lusophone countries. Moreover, although rapid advances in Natural Language Processing (NLP) and Artificial Intelligence (AI) have begun to reshape interpreting research and professional practice (Fantinuoli, 2018, 2023), the implications of such technologies for Chinese–Portuguese SI remain unexplored. Existing studies on computer-assisted interpreting tools, particularly automatic speech recognition (ASR) and real-time transcription, have focused predominantly on language pairs involving English (Defrancq & Fantinuoli, 2021; Pisani & Fantinuoli, 2021; Prandi, 2023; Yuan & Wang, 2023, 2024). No research has yet examined how these tools affect performance, interaction patterns, or interpreter perceptions in the Chinese–Portuguese combination.

These lacunae carry both practical and scholarly implications. As AI-powered tools become increasingly integrated into interpreting workflows, interpreters working with the Chinese–Portuguese pair face decisions about deploying such technologies under the demanding cognitive conditions of SI. Yet without empirical evidence on how real-time transcription affects performance, behavioral strategies, or output quality, practitioners and trainers lack a sound basis for evaluating technology adoption or designing pedagogical interventions.

To address these gaps, the present study investigates the integration of real-time automatic transcription in Chinese–Portuguese SI. The research pursues two objectives: (1) to assess the impact of ASR-generated transcription on interpreting performance, and (2) to examine how



interpreters interact with and perceive this technology during simultaneous task execution. These objectives are operationalized through the following research questions:

1. How does real-time automatic transcription affect the performance of interpreters in Chinese–Portuguese SI?
2. How do interpreters interact with and perceive real-time automatic transcription during the SI process?

Methodologically, the study employs a mixed-methods approach combining quantitative assessment of interpreting quality with qualitative data gathered through retrospective interviews and questionnaires. This design allows for triangulation across performance outcomes, behavioral patterns, and subjective experiences, providing a comprehensive account of technology integration in SI. By focusing on an underrepresented language pair and examining both the effects and experiential dimensions of ASR-assisted interpreting, this study makes three contributions. First, it expands empirical research on computer-assisted interpreting to the Chinese–Portuguese combination. Second, it deepens understanding of how interpreters negotiate real-time transcription as an additional information channel during cognitively demanding SI tasks. Third, it offers insights into the conditions under which human interpreters and AI technologies may coexist productively—a question of increasing relevance as the profession navigates rapid technological transformation.

2. Literature review

2.1 Technological developments and automatic speech recognition integration in simultaneous interpreting

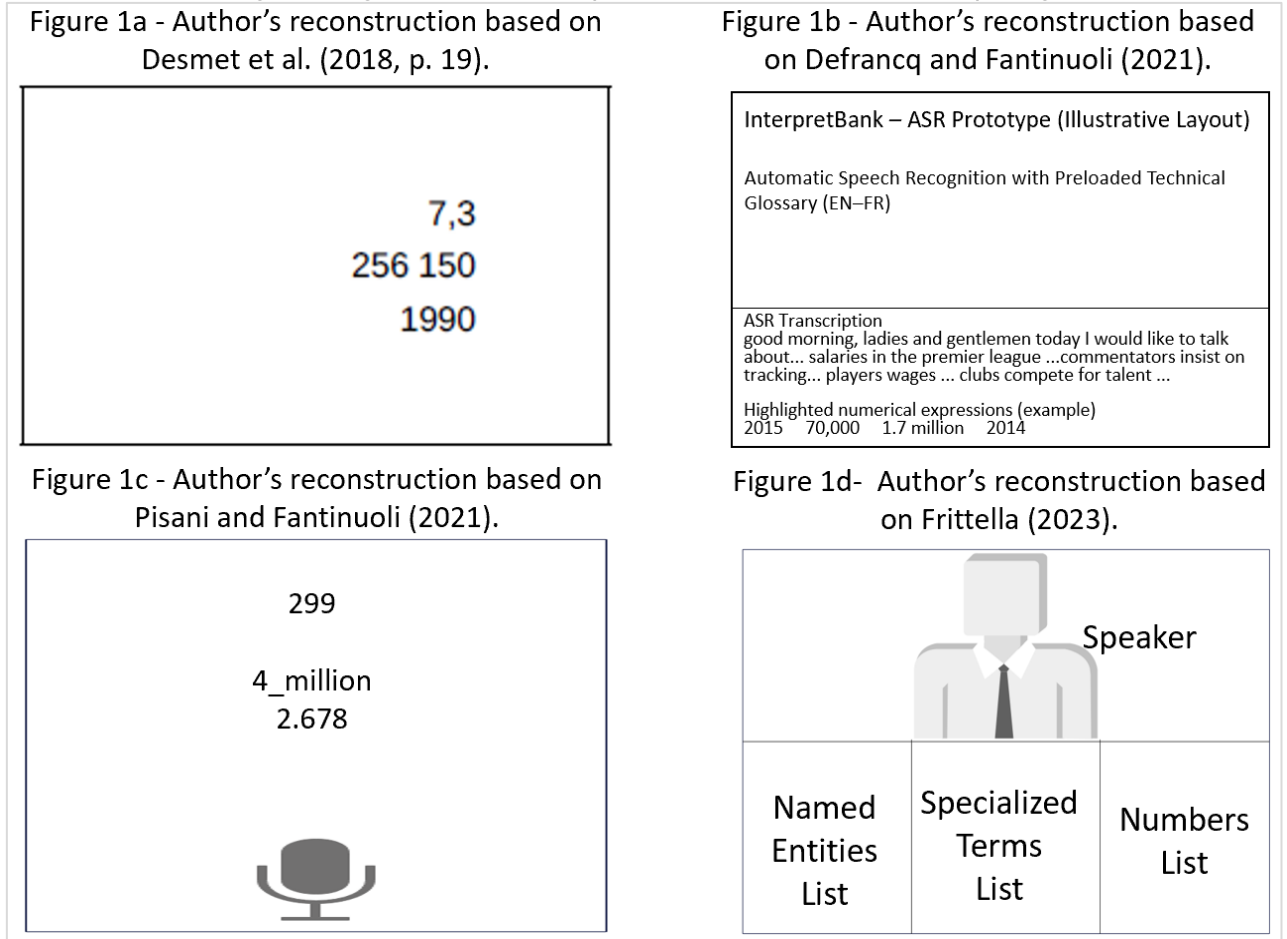
The integration of digital tools into interpreters' workflows has been a subject of scholarly inquiry for nearly three decades (Will, 2020). Early computer-assisted interpreting (CAI) tools consisted primarily of terminology management systems and digital glossaries designed to support lexical retrieval during SI. However, these first-generation tools required manual input and navigation, which imposed additional cognitive demands on interpreters already operating under considerable processing constraints. This limitation hindered widespread professional adoption, as the cost of attending to the tool often outweighed the potential benefits (Fantinuoli, 2018; Prandi, 2023).

A significant development came with the introduction of ASR-enabled systems, which, unlike earlier terminology tools requiring manual search and selection, can automatically detect and display relevant information from the speech stream, potentially reducing the need for interpreter-initiated queries (Fantinuoli, 2017). Desmet et al. (2018) provided early empirical evidence for this potential, examining ten MTI students performing SI tasks under two conditions: with and without technological support. To simulate ASR functionality, the researchers used pre-prepared PowerPoint slides displaying numerical content from the source speech as distinct visual units in real time (Figure 1a). Results indicated that technological support substantially improved number



rendition accuracy, from 56.5% without support to 86.5% with support—a reduction in errors by approximately two-thirds. The effect was particularly pronounced for complex numbers and decimals, categories typically associated with the highest error rates in SI.

Figure 1: Digital tools used in empirical studies of simultaneous interpreting



Source: Figure 1a is an author's reconstruction based on Desmet et al. (2018, p. 19); Figure 1b is an author's reconstruction based on Defrancq and Fantinuoli (2021); Figure 1c is an author's reconstruction based on Pisani and Fantinuoli (2021); Figure 1d is an author's reconstruction based on Frittella (2023).

Subsequent studies employing commercially available ASR-enabled tools have yielded broadly consistent findings. Defrancq and Fantinuoli (2021) examined the performance of six graduate-level interpreting students using InterpretBank, a terminology management tool incorporating ASR functionality (Figure 1b). With technological support, participants' accuracy in rendering numbers increased by 22.5 percentage points, a notable improvement though somewhat smaller than that reported by Desmet et al. (2018)—a difference potentially attributable to variations in experimental design, ASR accuracy, or participant characteristics. Pisani and Fantinuoli (2021) extended this line of inquiry in a larger-scale study involving 20 student interpreters, using InterpretBank integrated with Google Cloud's Speech-to-Text technology to automatically detect and display numerical content from the source speech (Figure 1c). Their analysis revealed that the technology-supported group exhibited a 14.8% error rate for numbers compared to 39.8% in the control condition, representing an average improvement of approximately 25 percentage points across all numerical

categories. Research has also examined the effects of CAI tools on other error-prone content categories. Frittella (2023) examined a CAI tool capable of automatically detecting and displaying numbers, proper names, and terminology from the source speech (Figure 1d), demonstrating improved accuracy across all three categories. These findings further reinforce the value of ASR-enabled support in addressing error-prone content in SI.

Beyond terminology-focused applications, more sophisticated ASR implementations have emerged in the form of real-time automatic transcription—also termed live transcription or streaming ASR. Unlike tools that extract and display only selected elements such as numbers or terminology, real-time transcription systems convert continuous speech into running text with minimal latency, generating output while the source speech is still in progress (He et al., 2019; Jurafsky & Martin, 2024). The underlying models are optimized to predict subsequent lexical items based on accumulated contextual information, resulting in near-instantaneous transcriptions of increasing accuracy. The evolution of these technologies has enabled their integration into a range of platforms, including videoconferencing software and specialized interpreting interfaces.

Several recent investigations have examined the application of real-time transcription in SI contexts. Cheung and Li (2022) conducted an experimental study with 24 students performing English-to-Chinese SI, comparing performance with and without access to ASR-generated live captions (Figure 2a).

Figure 2: Comparative presentation of the interfaces of automatic transcription tools

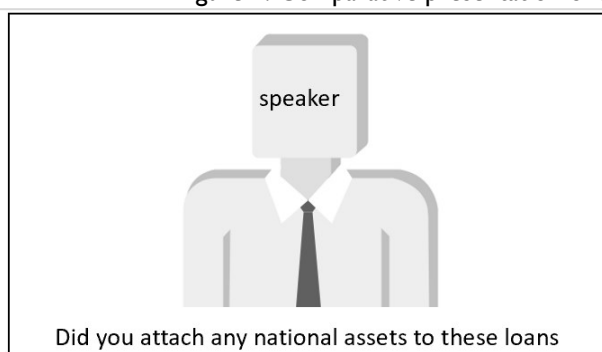


Figure 2a - Author's reconstruction based on Cheung and Li (2022, p. 8).

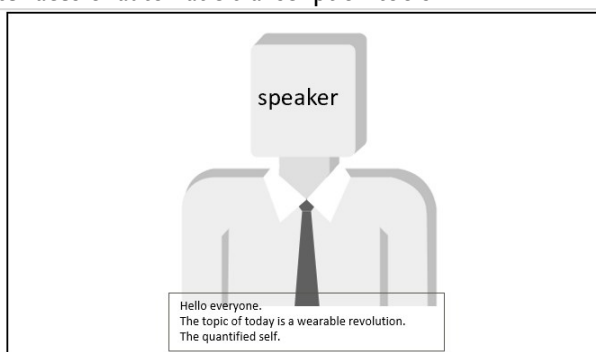


Figure 2b - Author's reconstruction based on Yuan and Wang (2024, p. 133).



Figure 2c - Authors' own screenshot: iflyrec captions in source language (font size 60)

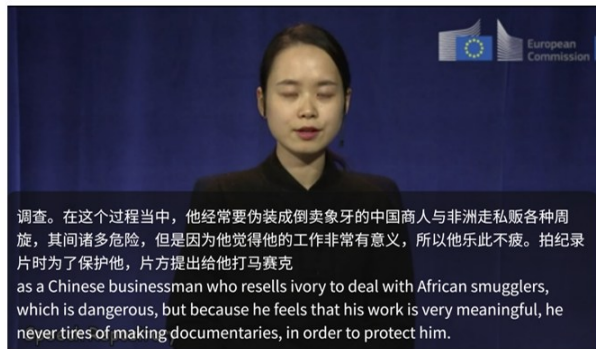


Figure 2d - Authors' own screenshot: iflyrec dual-language captions (font size 32)

Source: Figure 2a is an author's reconstruction based on Cheung and Li (2022, p. 8); Figure 2b is an author's reconstruction based on Yuan and Wang (2024, p. 133); Figures 2c and 2d are the authors' own screenshots of the iFlyrec interface used for illustrative purposes.

Results indicated significant improvements in rendition accuracy under the caption-supported condition, though a slight reduction in output fluency was observed—suggesting potential trade-offs between accuracy gains and delivery quality that warrant further investigation. Yuan and Wang (2023, 2024) investigated real-time automatic captions using the Zoom Meetings platform (Figure 2b), finding that transcription support reduced errors related to numbers and proper names. These findings reinforce the utility of ASR technologies for enhancing accuracy in factual details, while also highlighting the growing accessibility of such tools through mainstream communication platforms. Advances in interface design have further facilitated adoption by allowing transcription displays to be customized according to user preferences, including font size, positioning, and highlighting options (Figure 2c–2d). Such configurability may prove important for optimizing the balance between visual support and attentional demands during SI tasks.

Collectively, this body of research suggests that ASR and real-time transcription technologies may offer potential benefits for interpreters, particularly in reducing errors associated with numbers, proper names, and other high-density information. However, most existing studies have focused on language pairs involving English, predominantly English–German, English–French, and English–Chinese combinations. Research examining the effects of such tools on interpreter performance in Chinese–Portuguese SI is notably absent, despite the growing professional demand for services in this language combination.

2.2 Interpreter interaction with textual support and perceptions of technological tools in simultaneous interpreting

The increasing multimodality of interpreting environments has drawn scholarly attention to how interpreters engage with visual information during SI (Seeber, 2017). In professional settings, interpreters frequently encounter situations requiring them to process a spoken message while simultaneously consulting written materials—a configuration commonly termed “SI with text.” In such scenarios, interpreters must coordinate auditory input from the speaker with visual input from a source document, creating a dual-channel processing demand that requires careful allocation of attention and cognitive resources (Cammoun et al., 2009; Gile, 2009; Setton, 2015). Understanding how interpreters navigate this multimodal environment provides essential context for examining their interaction with newer forms of visual support, including ASR-generated transcription.

Empirical research on SI with text has consistently demonstrated interpreters’ heavy reliance on written materials when available. Cammoun et al. (2009), surveying 50 professional interpreters, found that 98% reported using textual support during SI with text, underscoring the perceived value of visual input in the interpreting process. Yang et al. (2020) examined how student interpreters managed SI with text under fast speech conditions in L1-to-L2 interpreting, finding that 75% of participants relied primarily on textual support while only 25% depended more heavily on auditory cues. This distribution suggests that visual input may serve a compensatory function when auditory processing becomes strained. However, such reliance carries risks. Chmiel et al. (2020) investigated professional interpreters working from L2 to L1 with texts containing intentionally incongruent items—instances where written content (proper names, numbers, and specific terms) diverged from the spoken message. Results revealed a striking disparity: accuracy reached 90% for congruent items



but dropped to 25% for incongruent items, indicating that interpreters may default to textual information even when it conflicts with auditory input.

Research employing eye-tracking methodology has further illuminated the temporal dynamics of visual attention during SI with text. Seeber et al. (2020) observed that professional interpreters' gaze typically lagged behind auditory processing, with fixations concentrated on textual segments preceding the current spoken content—a pattern suggesting that interpreters use text primarily to support output production. Zou et al. (2022), examining attention allocation among professional interpreters during L2-to-L1 SI with text, identified what they termed an “ear-leads-eye” pattern, wherein auditory input served as the primary guide while textual information provided supplementary support. These findings collectively suggest that experienced interpreters develop strategic approaches to managing dual-channel input, though the specific balance varies according to task demands and individual differences.

The introduction of dynamic technological tools—particularly real-time ASR transcription—raises new questions about how interpreters will manage visual support that updates continuously rather than remaining static. Unlike prepared scripts, ASR-generated text appears incrementally and may contain errors or unexpected formatting, potentially altering the interaction patterns observed in traditional SI with text. Several studies have begun investigating interpreter perceptions of such tools, with findings that highlight both opportunities and concerns.

Defrancq and Fantinuoli (2021), examining interpreter responses to the InterpretBank ASR prototype, reported that participants found the tool facilitative, particularly for managing numerical information displayed prominently within the transcription. However, interpreters also expressed concerns about the potentially distracting nature of constantly updating text and the risk of over-reliance on technological support even when output contained inaccuracies. Notably, participants disagreed regarding optimal display format: some preferred viewing only extracted numbers, while others favored joint presentation of numbers with surrounding context. Pisani and Fantinuoli (2021), investigating a configuration displaying numbers exclusively without full transcription, found moderate assessments of accuracy and clarity, with participants again citing the continuously present screen as a potential source of distraction.

Broader survey research has explored interpreter preferences for tool design. The Ergonomics for the Artificial Booth Mate project (EABM, 2021), involving 525 interpreters—predominantly experienced professionals—investigated preferences regarding the presentation of terminology, names, and numbers. Participants favored vertical arrangement with new items appearing below previous entries, and expressed preference for a unified display box over multiple columns, suggesting that visual simplicity facilitates integration into SI workflows. In the context of remote interpreting platforms, Saeed et al. (2022) conducted focus groups revealing that most participants preferred access to full speech transcripts, reinforcing the perceived value of comprehensive textual support.

Research specifically addressing real-time transcription has identified additional design considerations. Cheung and Li (2022) reported that interpreters found segmented subtitle presentation problematic, as fragmented display disrupted textual cohesion and complicated reading. This finding suggests that continuity of visual presentation may be important for effective integration of transcription support. Frittella (2023), interviewing 15 professional and academic interpreters,

found divergent preferences: some favored continuous full transcription while others preferred selective display of discrete elements such as numbers or terminology. Following practical experience with integrated ASR tools, participants emphasized the need for an adaptation period to become accustomed to the technology and identified potential attentional disruptions during use.

Collectively, this body of research indicates that interpreter interaction with visual support—whether traditional text or dynamic transcription—involves complex trade-offs between informational benefit and attentional cost. While interpreters generally value access to textual support, the specific format, update frequency, and content selection significantly influence usability and user experience. Importantly, most existing research has examined interpreter interaction and perceptions in experimental conditions involving predominantly English-paired combinations. How interpreters working with the Chinese–Portuguese pair engage with and perceive real-time transcription tools during SI remains unexplored, representing a gap that the present study addresses.

3. Methodology

3.1 Profile and selection of participants

The study recruited native Mandarin Chinese speakers enrolled in the first year of master's programs in Portuguese-language studies at universities in mainland China, Macau, and Portugal. This population was selected on the basis that such students would possess higher Portuguese proficiency than undergraduates while remaining at an early stage of SI training. We anticipated that participants without extensive professional experience would exhibit responses to technological support more indicative of initial adaptation processes, uninfluenced by established habits or preconceptions that experienced interpreters may have developed.

A recruitment poster outlining the study's purpose and requirements was disseminated through WeChat groups for Chinese students in Portuguese-language programs, including groups for students studying abroad and domestic university cohorts, reaching an estimated audience of over 2,000 individuals. Participants received only modest reimbursement for time and incidental expenses related to participation; no research-related fees were charged, in line with national ethical principles on participant compensation (National Health Commission of the People's Republic of China [NHC], 2023, Art. 17(4)). This recruitment strategy yielded a final sample of ten participants who met the inclusion criteria and completed all experimental tasks.

The participant profile was predominantly female (nine of ten). All participants were aged 18–25 years and enrolled in the first year of master's programs focusing on Portuguese language or translation studies. Regarding institutional affiliation, seven participants were enrolled at universities in Macau: four at Macau Polytechnic University and three at the University of Macau. Two participants were from Portugal, both enrolled at the University of Lisbon. One participant was from mainland China, enrolled at Beijing Foreign Studies University. Most participants reported limited prior interpreting experience and minimal familiarity with digital tools for interpreting.

The study followed institutional principles on research integrity and participant protection that are now formalized in the *Macao Polytechnic University Regulations on Research Ethics and Integrity*



(effective July 4, 2025), together with institutional guidance on informed consent and ethical procedures for investigations involving human participants. All participants provided voluntary written informed consent prior to participation. They were informed of the study's purpose, procedures, audio recording, data handling, and their right to withdraw at any time without penalty. The study involved minimal-risk behavioral tasks performed by adult volunteers, collected no sensitive personal data, and used no deception. In addition, the study observed national ethical principles regarding written informed consent and reasonable participant compensation (NHC, 2023, Arts. 17(4) and 33).

3.2 Materials and tools

Source speech materials were drawn from the Speech Repository of the European Commission's Directorate-General for Interpretation¹, a publicly accessible database of interpreting training materials. Speeches in this repository are classified by language, difficulty level, thematic domain, and intended pedagogical use, with regular evaluation by professional interpreters and experienced trainers ensuring quality and relevance. Given the participants' early-stage training, speeches designated as beginner level—the lowest difficulty classification available—were selected. Restricting the search to Chinese-language speeches in the general domain yielded 18 candidate items, from which ten covering diverse topics—including international event organization, budget travel, consumer behavior, health and lifestyle, and technology platforms—were selected for further evaluation.

To validate speech selection and ensure comparable difficulty across experimental conditions, a difficulty rating survey was conducted using the Wenjuanxing² online platform. Seventeen raters participated, comprising eight professional interpreters, three interpreter trainers, and six interpreter trainees (nine female, eight male). Raters assessed each speech's SI difficulty on a scale of 0–10. The objective was to identify two speeches of equivalent difficulty: one for the SI condition without transcription support, and one for the SI condition with real-time transcription.

Survey results identified two speeches with closely matched difficulty ratings: 5.94 and 5.59, respectively. Shapiro-Wilk tests confirmed that both rating distributions satisfied normality assumptions ($p > .05$), permitting parametric comparison. A paired-samples t-test revealed no statistically significant difference between the two difficulty ratings ($t = 1.54$, $p = .144$), supporting their equivalence for experimental purposes. The first speech, on “China's experience hosting international events”, was assigned to the SI condition without transcription support; the second, on “the founding of the travel platform Qyer.com”, was assigned to the SI condition with real-time transcription.

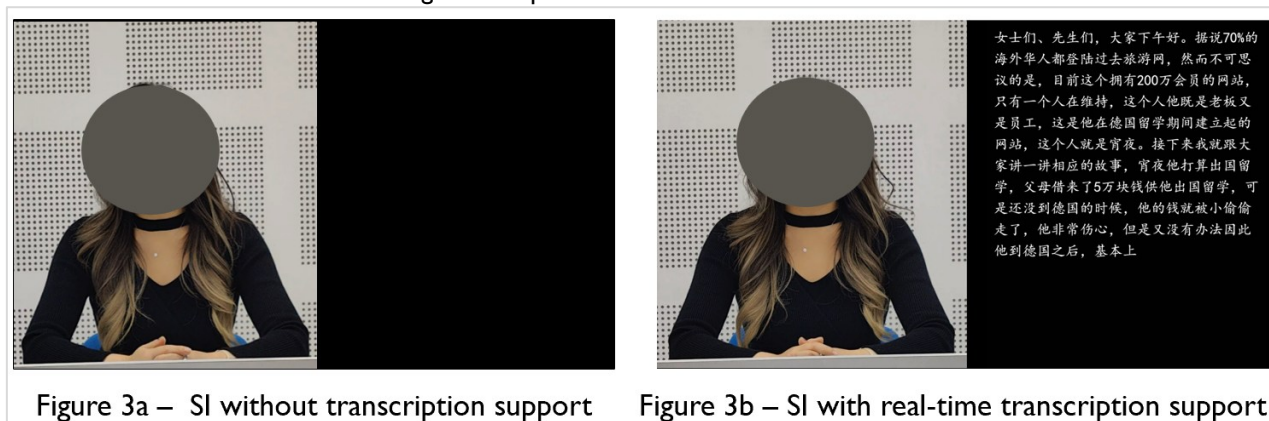
To ensure stimulus consistency, a female native Mandarin speaker was recorded reading both speeches at an even pace without accompanying gestures. Videos were edited to approximately four minutes' duration, comprising approximately 730 characters each. A standardized visual interface was developed with two sections: speaker video on the left, and a designated display area on the right. In the condition without transcription support, this area displayed a black background; in the

¹ Available at: <https://speech-repository.webcloud.ec.europa.eu/>.

² Available at: <https://www.wjx.cn/>.

condition with transcription support, real-time Chinese transcription appeared in white STKaiti font (size 30) with 1.5 line spacing. Figure 3 illustrates both interface configurations.

Figure 3: Experimental interfaces for SI tasks



Source: The Authors.

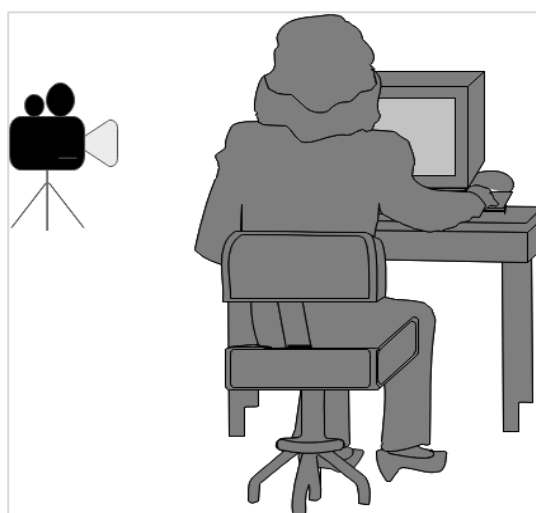
Real-time transcription was generated using iFlytek's iFlyrec³, an ASR platform supporting multiple languages with low-latency output. According to official specifications, iFlyrec achieves a Word Error Rate (WER) of 2% and average latency of 0.2 seconds. Performance was verified during pilot testing, yielding a measured WER of 3.60% and sentence-level latency of 1.24 seconds—slightly worse than the vendor specifications but acceptable for experimental purposes. Overall, iFlyrec provided reliable, low-error transcription suitable for evaluating the impact of ASR support on SI performance.

3.3 Experimental procedures

The experimental protocol comprised two phases: a preparatory phase and a main experimental session. The preparatory phase spanned the four consecutive days preceding the main experiment, during which participants completed daily interpreting exercises of approximately five minutes each. To ensure uniform preparation and monitor protocol compliance, participants recorded their practice sessions using smartphones while viewing materials displayed on their computer screens. Recordings were submitted to the researchers upon completion of each daily exercise. Figure 4 illustrates the participant self-recording setup during this phase.

The main experimental session was conducted remotely via Tencent Meeting, an online conferencing

Figure 4: Participant self-recording setup during preparatory exercises



Source: The Authors.

³ Available at: <https://www.iflyrec.com/>.

platform accessible to all participants. One day prior to the session, participants received meeting access credentials and detailed procedural instructions. Upon joining the session, video recording commenced automatically to document the proceedings. Before each interpreting task, participants received a glossary via the platform's chat function and were permitted up to five minutes for preparation, after which source speeches were presented via screen sharing. The experimental session comprised multiple interpreting conditions; the present analysis focuses on two conditions directly relevant to the research questions: SI without transcription support and SI with real-time transcription support. To counterbalance potential order effects, task presentation order was randomized across participants, and breaks were provided between tasks upon request to mitigate fatigue.

Following completion of the interpreting tasks, participants completed a post-experiment questionnaire addressing their perceptions and experiences with the technological support. Additional process data were collected during the session; however, given the present study's focus on performance outcomes and interpreter perceptions of ASR support, these data are reserved for future analysis.

3.4 Data collection and analysis

Interpreter performance was assessed by a team of five raters, all female, aged 25–30, and currently enrolled in doctoral programs in Portuguese Studies. Each rater possessed several years of formal Portuguese language training, with Mandarin Chinese as their native language, ensuring competence in both source and target languages. These qualifications were essential for evaluating interpretations from Chinese to Portuguese with sensitivity to both linguistic accuracy and target language quality.

The assessment employed the scale developed by Han (2021), which has demonstrated validity in prior SI research (Han, 2015a, 2015b, 2018). This scale comprises four dimensions rated on an 8-point scale, which served as the evaluation criteria: information completeness, assessing accuracy and thoroughness of message transmission; fluency, examining the presence of disfluency markers; target language quality, evaluating idiomatic appropriateness, naturalness, and grammatical accuracy; and holistic performance, providing an integrated assessment of overall interpreting effectiveness. To establish rating consistency, the process commenced with independent evaluation of four randomly selected interpretation recordings by all raters using these criteria. A subsequent calibration session of approximately 90 minutes enabled collective discussion and alignment of scores across scale dimensions, standardizing application of the descriptors. Following calibration, raters were divided into two groups of three and two members respectively to distribute the workload, with each interpretation evaluated by two raters to ensure cross-verification. To triangulate performance data, Ear-Voice Span (EVS)—defined as the temporal interval between the onset of source speech segments and the onset of corresponding target language output (Guo & Han, 2024)—was also measured.

In addition to performance measures, qualitative data were collected through questionnaires and retrospective protocol interviews. The post-experiment questionnaire comprised 20 items designed to capture participants' perceptions of task difficulty, attention allocation strategies, and

evaluations of the transcription tool's utility (see Appendix A). The present analysis draws primarily on items directly relevant to the research questions; additional data were collected for future investigation. Items employed various response formats including 7-point and 5-point Likert scales, semantic differentials, and multiple-choice questions, and drew upon established instruments including the System Usability Scale (Brooke, 1996), the Computer System Usability Questionnaire (Lewis, 1995), the User Experience Questionnaire (Laugwitz et al., 2008), and the Technology Acceptance Model (Davis, 1989). Following task completion, participants engaged in retrospective protocol interviews in which they reflected on their experiences, including their interaction with the transcription display and decision-making processes during interpretation. The interview protocol addressed perceptions of task difficulty, the role of visual-verbal input in interpreting strategy, and recommendations for tool improvement (see Appendix B).

Quantitative analysis began with calculation of descriptive statistics including means and standard deviations. Data normality was assessed using Shapiro-Wilk tests; for normally distributed data, paired-samples t-tests were employed to compare performance across conditions, while Wilcoxon signed-rank tests served as the nonparametric alternative for non-normally distributed data. Interview recordings were transcribed using Tencent's transcription service, after which transcripts were analyzed thematically using Microsoft Word and Excel to identify emergent patterns regarding participants' perceptions of the transcription tool, interface preferences, and interpreting strategies.

4. Results

4.1 Effects of real-time transcription on interpreting performance

Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC). The two-way mixed model with absolute agreement, ICC (3,1), was adopted to evaluate consistency across the four performance dimensions. The ICC values obtained were 0.751 for information completeness, 0.698 for fluency, 0.636 for target language quality, and 0.741 for holistic performance.

Table 1: Mean performance scores by condition and dimension

Participant	Completeness		Fluency		Quality		Holistic	
	with	without	with	without	with	without	with	without
1	4.0	3.0	4.5	3.0	3.5	3.5	4.0	3.0
2	5.0	4.5	4.5	3.5	4.5	4.0	4.5	4.5
3	2.5	1.5	2.0	1.0	2.0	2.0	2.5	1.5
4	2.0	2.0	1.5	2.0	1.5	2.0	1.5	2.0
5	4.5	3.5	4.5	4.0	3.5	3.5	4.5	4.0
6	3.5	3.0	2.0	3.0	2.5	3.0	3.0	3.0
7	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
8	1.5	1.5	2.0	1.5	2.0	2.5	2.0	1.5
9	1.5	1.0	2.0	1.0	1.5	2.0	1.5	1.0
10	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Average	3.0	2.5	2.8	2.4	2.6	2.8	2.9	2.6
Standard deviation	1.4	1.3	1.4	1.2	1.2	1.0	1.3	1.3

Source: The Authors.



According to Koo and Li (2016), ICC values between 0.50 and 0.75 indicate moderate reliability, while values between 0.75 and 0.90 indicate good reliability. The present values thus fall within the moderate to good range, supporting the use of averaged ratings in subsequent analyses. Table 1 presents individual and mean scores for all participants across both experimental conditions. In the condition with real-time transcription, mean scores were 3.0 for information completeness, 2.8 for fluency, 2.6 for target language quality, and 2.9 for holistic performance. In the condition without transcription, corresponding scores were 2.5, 2.4, 2.8, and 2.6. Scores were higher with transcription support in three of four dimensions, while target language quality showed a slight, non-significant decrease.

Table 2: Paired t-test results comparing conditions

Comparison	t	p	Cohen's d _{av}
Completeness	3.25	.010*	0.33
Fluency	1.63	.137	0.30
Quality	-1.41	.193	-0.14
Holistic	1.96	.081	0.23

Source: The Authors.

Note: *p < .05. d_{av} values are standardized mean differences based on the average standard deviation of the two paired conditions.

Paired t-tests were conducted following verification of normality assumptions using the Shapiro-Wilk test. As shown in Table 2, a statistically significant difference was found only for information completeness (t = 3.25, p = .010, d = 0.33). The remaining dimensions showed no significant differences: fluency (t = 1.63, p = .137), target language quality (t = -1.41, p = .193), and holistic performance (t = 1.96, p = .081).

To contextualize the quantitative results, we further examined representative segments from participants' output under both conditions.

Example 1: simultaneous interpreting without transcription support

Source text (Chinese): 很多抗议 G 20 的、持不同意见的人士。那么他们在汉堡砸了很多商店。然后还烧了很多路边的汽车。造成了整个汉堡市非常混乱的一种情况。

Reference interpretation: Muitos manifestantes contra o G20, que tinham opiniões divergentes, vandalizaram várias lojas e incendiaram carros nas ruas, causando uma situação de caos generalizado na cidade de Hamburgo.

- Participant 2: Existem muitas manifestações, então elas vão dar as lojas de hambúrguer e prejudicar os carros e causando a difusão.
- Participant 5: Tem muitas manifestações e elas vandalizaram muitas lojas e destruíram muitos carros. Toda esta faz que um caos.
- Participant 7: Há muitas manifestações durante o evento. Vandalizaram muitas lojas em Hambúrguer. Os carros levam à confusão no tráfego.



- Participant 10: Por isso vandalizam lojas e causar a cidade de hambúrguer um grande problema.

In this example, participants encountered multiple difficulties in maintaining completeness and semantic accuracy. Several participants confused “汉堡” (Hànǎo, Hamburg) with “汉堡包” (Hànǎobāo, hamburger), as these words share the same initial characters in Chinese. This confusion resulted in semantic distortion, with participants rendering the city name as “hambúrguer” or “lojas de hambúrguer.” Additionally, key contextual elements such as the reference to the G20 summit (“G20”) and the specific location (“汉堡市”, Hamburg) were omitted by most participants. While Participant 5 conveyed the general idea of “caos,” critical details were lost, reducing overall information completeness. The grammatical errors in participants’ output (e.g., “muitos manifestações,” “causando a difusão”) reflect the cognitive demands of the task and have been preserved as produced.

Example 2: simultaneous interpreting with real-time transcription support

Source text (Chinese): 这是一家四星级的酒店，但是他打出的广告却是一欧元。当时肖毅觉得有点不可思议，但是抱着试试看的心态去试了试。结果，他真的就花了 1 欧元住了一晚四星级的酒店。

Reference interpretation: Este é um hotel de quatro estrelas que anunciou uma oferta de um euro por noite. Xiao Yi achou isso incrível, mas decidiu experimentar. E, surpreendentemente, ele realmente conseguiu passar a noite por apenas um euro.

- Participant 1: Viu a promoção do hotel da luxo, mas a publicidade é um euro. Xiao Ye acha que não era possível, mas queria tentar e ele realmente pagou um euro para viver neste hotel.
- Participant 5: Ela viu, tem uma publicação sobre este hotel é que só precisa de um euro. Por isso quer tentar este hotel.
- Participant 8: Xiao Ye viu um hotel de luxo na Suíça de quatro estrelas.
- Participant 10: Aquela altura. Ele acha que não é possível. Mas ela tenta. Uma vez e realmente ele...

In this condition, Participant 1 captured the essential narrative with reasonable completeness, despite some simplification. However, the ASR system incorrectly transcribed the name “肖毅” (Xiāo Yì) as “宵夜” (Xiāo Yè), and several participants reproduced this error as “Xiāo Yè” in their output. This instance illustrates a specific limitation of transcription support: when ASR errors occur, especially in proper names, interpreters may inadvertently reproduce them.



In summary, the quantitative analysis indicates that real-time transcription was associated with significantly higher scores in information completeness, while differences in other dimensions did not reach statistical significance. The qualitative examples suggest that transcription support may help interpreters capture content that might otherwise be missed, but also reveal the risk of error propagation when ASR accuracy is compromised.

4.2 Interpreter interaction with transcription display

To investigate how interpreters interacted with real-time transcription during SI, data were collected through post-experiment questionnaires and retrospective interviews. Participants were asked to estimate the proportion of attention allocated to the transcription display versus the audio input during the task.

Table 3: Self-reported attention allocation in simultaneous interpreting with real-time transcription

Participant	Transcription information (%)	Auditory information (%)
1	84	16
2	94	6
3	85	15
4	30	70
5	65	35
6	60	40
7	64	36
8	49	51
9	53	47
10	20	80
Average	60.40	39.60
Standard deviation	23.73	23.73

Source: The Authors.

As shown in Table 3, participants reported allocating an average of 60.40% of their attention to the transcription display (SD = 23.73), with the remaining 39.60% directed toward auditory input. However, substantial individual variation was observed (SD = 23.73), indicating heterogeneous interaction strategies: two participants (P4 and P10) reported allocating 70% or more attention to audio, while three others (P1, P2, and P3) reported allocating over 80% to the transcription. This distribution suggests that interaction with the tool was predominantly reading-oriented for most participants, though individual strategies varied substantially.

Retrospective interviews provided further insight into attention allocation patterns. Several participants reported that, at the beginning of the task, they were able to distribute attention relatively evenly between auditory and visual channels. However, as the task progressed and the pace of input increased, attention shifted progressively toward the transcription display. Participants explained that when information flow exceeded their processing capacity, they relied more heavily on the visual text to avoid falling too far behind the speaker. Several participants described the experience as functionally closer to sight translation than to traditional SI, as they found themselves reading from the screen rather than processing audio in real time.

Only two participants reported exercising conscious control over attention allocation, highlighting the largely automatic and visually driven nature of attention shifts when text is present.



One noted that, despite intending to prioritize the auditory channel, attention frequently shifted to the transcribed text almost involuntarily. This suggests a natural tendency toward visual processing when text is available, which may have implications for interpreting strategy and performance.

The shift toward visual processing was also reflected in EVS measurements. Table 4 presents mean EVS values for both experimental conditions.

Table 4: Mean ear-voice span by condition (seconds)

Participant	SI with transcription	SI without transcription
1	19.51	13.00
2	19.37	5.87
3	26.35	10.74
4	24.39	12.75
5	19.69	8.14
6	18.13	8.15
7	12.89	9.78
8	8.33	4.38
9	8.32	4.74
10	4.92	3.90
Average	16.19	8.15
Standard deviation	7.22	3.39

Source: The Authors.

The mean EVS in the transcription condition was 16.19 seconds (SD = 7.22), compared to 8.15 seconds (SD = 3.39) in the condition without transcription. EVS values were calculated using segment onset alignment, consistent with standard procedures (Guo & Han, 2024). This difference suggests that when transcription was available, interpreters tended to wait for more text to accumulate on screen before producing output, resulting in greater temporal distance from the source speech. Several participants confirmed this pattern in interviews, noting that they often waited for complete sentences to appear on screen before beginning to interpret, rather than following the speaker’s delivery in real time.

In summary, the interaction data indicate that most interpreters adopted a reading-oriented strategy when transcription was available, allocating the majority of attention to visual input. This pattern was associated with longer EVS values, suggesting a shift from real-time audio processing toward a more text-dependent approach.

4.3 Perceptions and preferences regarding transcription support

Participants’ perceptions of the transcription tool were assessed through standardized usability measures and retrospective interviews. Table 5 summarizes scores on four metrics: the SUS, the UEQ, the TAM, and the CSUQ. The mean SUS score of 56.25 falls below the moderate perceived usability of 68, indicating limited perceived usability and suggesting that the system may require further refinement. The UEQ score of 0.73 (scale: –3 to +3) suggests a mildly positive user experience, though with considerable variation across participants (range: –0.75 to 2.38). The TAM score of 3.66 (scale: 1–5) reflects moderate technology acceptance. CSUQ subscale scores for information quality (4.32) and interface quality (4.68) on a 7-point scale indicate generally favorable perceptions, though not without reservations.

Table 5: User experience metrics for the transcription tool

Participant	SUS	UEQ	TAM	Quality of Information (CSUQ)	Interface Quality (CSUQ)
1	32.50	-0.50	3.63	2.00	3.25
2	67.50	1.50	4.88	6.00	6.00
3	60.00	-0.38	2.25	4.00	3.75
4	40.00	0.75	4.00	5.00	5.00
5	37.50	1.50	3.50	4.50	5.25
6	75.00	0.88	3.38	5.33	3.75
7	77.50	1.38	4.00	5.33	6.00
8	52.50	0.50	3.13	2.50	5.25
9	77.50	2.38	4.50	6.17	6.00
10	42.50	-0.75	3.38	2.33	2.50
Average	56.25	0.73	3.66	4.32	4.68
Standard deviation	17.57	1.02	0.74	1.55	1.27

Source: The Authors.

Interview data revealed generally positive—but nuanced—perceptions of the transcription tool. Eight of ten participants viewed real-time transcription as beneficial to their performance rather than as a distraction. Participants identified several specific advantages. First, the transcription provided a visual roadmap of content, helping them anticipate upcoming information and organize their output more effectively. This was considered particularly valuable given the syntactic differences between Chinese and Portuguese, which often require restructuring during interpretation. Second, participants noted that visual confirmation of auditory input reduced uncertainty and working memory load, allowing them to focus on reformulation rather than retention.

However, participants also noted several challenges associated with transcription use, particularly related to accuracy and cognitive load. Several noted that monitoring both audio and text was cognitively demanding, leading to increased fatigue compared to traditional SI. Transcription accuracy emerged as a critical concern: recognition errors could disrupt the flow of interpretation and, as illustrated in Example 2, lead to error propagation in output. Additionally, some participants observed that the availability of complete text on screen encouraged them to convey more information than they would typically attempt in SI without visual support, which sometimes caused them to fall further behind the speaker.

Regarding tool preferences, nine of ten participants indicated willingness to use transcription support in future interpreting work. Preferences regarding device type varied: most participants preferred using a computer, while a smaller number favored tablets or smartphones. Seven participants expressed a preference for transcription displayed in a separate window within the booth, rather than integrated with video. For display format in Chinese-to-Portuguese interpretation, preferences diverged: six participants preferred source language text with problem indicators (numbers, names, terminology) highlighted in the target language, two preferred full bilingual display, one preferred target language only, and one suggested displaying only problem indicators bilingually.

In summary, interpreters' perceptions of real-time transcription were generally positive, with most recognizing its potential to reduce cognitive load and support information capture. At the

same time, concerns about transcription accuracy, cognitive fatigue, and the risk of over-reliance point to the need for continued refinement of such tools, particularly for language pairs with substantial structural differences.

5. Discussion and implications

5.1 Effects of transcription on performance and implications for practice

Quantitative results show a significant positive effect of real-time transcription on information completeness, with no significant differences in fluency, target language quality, or holistic performance. This pattern suggests that transcription support offers selective rather than comprehensive benefits to interpreting performance.

The improvement in information completeness aligns with previous research demonstrating that visual support can enhance accuracy in SI, particularly for problem triggers such as numbers and proper nouns (Cheung & Li, 2022; Yuan & Wang, 2023, 2024). Our results extend these findings to the Chinese–Portuguese language pair. Interview data corroborated this interpretation: participants reported feeling more confident in capturing source content when transcription was available, as the visual display allowed them to verify information that might otherwise be missed. This benefit is consistent with Guo et al. (2025a), who found that SI with automatic live captioning yielded higher quality than SI without text, though not matching the level achieved with prepared text materials.

The absence of significant improvement in other performance dimensions, particularly target language quality, merits attention. Two factors may account for this pattern. First, research on SI with text has documented a tendency for interpreters to default to visual input, which can result in output that follows source language structures more closely, potentially compromising target language idiomaticity (Chmiel et al., 2020; Gile, 2009; Ma & Cheung, 2020; Shreve et al., 2010). Second, and perhaps more critically in the ASR context, transcription errors can propagate into interpreter output. As illustrated in Section 4.1, several participants reproduced ASR misrecognitions (e.g., “Xiao Ye” for “Xiao Yi”), suggesting that the visual channel may override auditory information even when the latter is more accurate. These findings highlight a fundamental tension in ASR-assisted SI: the same visual support that aids information capture may also introduce new sources of error.

The EVS data provide additional insight into how transcription altered interpreting behavior. The substantially longer EVS in the transcription condition (16.19 vs. 8.15 seconds) suggests that interpreters shifted from real-time listening toward a reading-and-reformulating approach, waiting for text to accumulate before producing output. While this strategy may facilitate information capture, it distances the interpreter from the temporal flow of source speech, representing a trade-off between completeness and immediacy. This shift has direct implications for how interpreters balance immediacy and completeness in time-sensitive SI tasks.

These findings carry implications for interpreter training. As transcription tools become more widely adopted, training programs will need to address the competencies required for their effective use. Recent research supports a dual-focused approach that builds traditional SI skills while incorporating specialized modules for technology-assisted interpreting, as performance in



conventional SI modes has been shown to transfer to ASR-assisted contexts (Guo et al., 2025b). Our results reinforce this direction: trainees need explicit instruction in attention allocation strategies, learning to use transcription as a supplement to rather than a replacement for auditory processing. Training might include exercises with imperfect transcription to help interpreters calibrate their reliance on visual input, recognizing when to trust and when to override the displayed text.

In summary, real-time transcription offers genuine benefits for information capture while introducing challenges related to attention management and error propagation. Realizing the potential of this technology will require both continued refinement of ASR accuracy and pedagogical approaches that prepare interpreters to integrate visual and auditory input effectively.

5.2 Interaction patterns, user perceptions, and implications for tool design

The second research question addressed how interpreters interact with and perceive real-time transcription. Beyond documenting interaction patterns and user evaluations, our findings reveal specific preferences regarding tool design that carry implications for technology development.

The interaction data indicate that most participants adopted a reading-oriented strategy, allocating the majority of attention to the transcription display. This visual orientation intensified as tasks progressed: when information flow exceeded processing capacity, participants increasingly relied on the textual display, with some describing the experience as resembling sight translation. This pattern is consistent with research on SI with text (Chmiel et al., 2020; Wang et al., 2025) and suggests that interpreters naturally gravitate toward visual input when available. The extended EVS observed in the transcription condition reflects this behavioral shift, as interpreters waited for text to accumulate before producing output. User perceptions were generally positive, with most participants viewing transcription as beneficial for information capture and working memory relief, though concerns about cognitive fatigue and ASR accuracy were also noted.

Regarding interface preferences, our study broadened the scope of investigation beyond the typical focus on terminology management tools. Participants were exposed to various display formats, including monolingual and bilingual transcriptions as well as terminology suggestion systems. A clear trend emerged: most interpreters preferred source language transcription with problem triggers (numbers, terminology, proper nouns) highlighted alongside their target language equivalents. This preference for comprehensive display aligns with Saeed et al. (2022), whose focus groups indicated strong preference for platforms that included full transcripts, and with Cheung and Li (2022), who found that interpreters preferred complete sentences over fragments.

With regard to layout, our findings accord with the EABM project (2021), which concluded that most users prefer vertical arrangement where new content appears below previous text. However, preferences were not uniform across all dimensions. Frittella (2023) documented divided preferences between full transcriptions and segmented displays. Our findings and those of Guo et al. (2025c) both show strong preference for continuous transcription over problem-trigger-only displays. Guo et al. (2025c) suggest that isolated display of problem triggers may disrupt cognitive flow, whereas continuous transcription supports more cohesive processing. However, our participants showed greater preference for bilingual display with highlighted problem triggers, while

Guo et al.'s (2025c) participants favored source language transcription with target language captions. This difference may reflect the distinct demands of different language pairs: the substantial syntactic divergence between Chinese and Portuguese may increase interpreters' reliance on target language support for reformulation.

These findings point to customization as a central design priority. The diversity of preferences observed, both within our study and across the literature, suggests that no single display format will suit all interpreters. Effective tools should allow users to adjust information density, display format, and visual presentation according to individual needs. Additionally, given the documented concerns about ASR accuracy and its direct impact on output quality, continued refinement of recognition performance remains essential, particularly for problem triggers. The development of transcription tools for SI should proceed through ongoing dialogue between technologists and end users, ensuring that design decisions are informed by the practical realities of interpreting work.

In summary, interpreters demonstrated clear preferences for comprehensive, continuous transcription displays while exhibiting substantial individual variation. These findings underscore the need for flexible, customizable interfaces that accommodate diverse processing styles and support effective integration of transcription technology into SI practice.

6. Conclusion

This study investigated the integration of real-time automatic transcription in Chinese–Portuguese SI, addressing two research questions concerning performance effects and interpreter interaction with the technology.

Regarding performance, the findings indicate that transcription support significantly improved information completeness while producing no significant effects on fluency, target language quality, or holistic performance. This selective benefit suggests that ASR-generated transcription functions primarily as a safety net for content capture rather than as a comprehensive performance enhancer. The qualitative analysis revealed a corresponding risk: interpreters may reproduce ASR errors, particularly in proper names, when visual input overrides auditory processing. Regarding interaction and perceptions, self-reports and EVS patterns suggest predominantly reading-oriented strategies, with attention increasingly directed toward the transcription display as task demands intensified. This shift was reflected in substantially longer EVS, indicating a move from real-time listening toward text-dependent processing. Despite moderate usability ratings, most participants viewed the technology favorably and expressed clear preferences for comprehensive, continuous displays with highlighted problem triggers—preferences that diverged in certain respects from findings in other language pairs, potentially reflecting the specific demands of Chinese–Portuguese interpreting.

These findings contribute to the literature in three respects. First, they extend empirical research on ASR-assisted interpreting to an underrepresented language pair, demonstrating that patterns observed in English-paired combinations partially generalize to the Chinese–Portuguese context while also revealing language-pair-specific preferences. Second, the study provides evidence for the dual nature of transcription support—beneficial for information capture yet potentially detrimental when ASR accuracy falters—highlighting the importance of error management in both tool design and interpreter training. Third, the documented individual variation in interaction



patterns and preferences reinforces the need for customizable interfaces that accommodate diverse processing styles.

Several limitations warrant acknowledgment. The sample of ten first-year master's students, while appropriate for an exploratory study, limits generalizability to professional interpreters or other training levels. The use of self-reported attention allocation, rather than objective measures such as eye-tracking, provides only an approximate account of visual processing behavior. The two source speeches, though matched for difficulty, represented a limited range of content types and speaking styles. Additionally, the cross-sectional design captured initial responses to transcription support but could not assess how interaction patterns might evolve with extended exposure and practice.

Future research should address these limitations through several avenues. Studies incorporating eye-tracking methodology would provide more precise data on visual attention allocation and reading patterns during ASR-assisted SI. Longitudinal designs could examine how interpreters adapt to transcription support over time and whether the observed tendency toward visual over-reliance diminishes with experience. Comparative studies across multiple language pairs would clarify which findings reflect universal patterns of human-technology interaction and which are specific to particular linguistic combinations. Finally, research examining professional interpreters alongside trainees would illuminate how expertise moderates the effects of transcription support on performance and strategy.

In conclusion, real-time automatic transcription represents a potentially valuable but not unproblematic addition to the interpreter's toolkit. The technology offers genuine support for information capture while simultaneously introducing risks of error propagation and visual over-reliance. Realizing its potential will require coordinated efforts: continued improvement in ASR accuracy, particularly for problem triggers; interface designs that support flexible, individualized use; and training approaches that prepare interpreters to integrate visual and auditory input strategically. As AI-powered tools become increasingly embedded in interpreting workflows, empirical research of this kind provides an essential foundation for evidence-based decisions by practitioners, trainers, and technology developers alike.

References

- Amaro, V., & Pires, M. J. (2024). Found in Translation, Lost in Education: Artificial Intelligence's Impacts on Translation Tertiary Education in Macao. *Asian Education and Development Studies*, 13(4), 269–281. <https://doi.org/10.1108/AEDS-01-2024-0012>
- Brooke, J. (1996). SUS: A 'Quick and Dirty' Usability Scale. In P. W. Jordan, B. Thomas, I. L. McClelland & B. Weerdmeester (Eds.), *Usability Evaluation in Industry* (pp. 189–194). CRC Press.
- Cammoun, R., Davies, C., Ivanov, K., & Naimushin, B. (2009). Simultaneous Interpretation with Text: Is the Text 'Friend' or 'Foe'? Laying Foundations for a Teaching Module [Seminar paper]. University of Geneva.
- Chao, L. S., Wong, D. F., Ao, C. H., & Leal, A. L. (2018). UM-PCorpus: A Large Portuguese-Chinese Parallel Corpus. In E. Yang & L. Sun (Eds.), *Proceedings of the LREC2018 Workshop "Belt &*



Road: *Language Resources and Evaluation*” (pp. 38–43). Beijing Advanced Innovation Center for Language Resources.

- Cheung, A. K. F., & Li, T. (2022). Machine Aided Interpreting: An Experiment of Automatic Speech Recognition in Simultaneous Interpreting. *Translation Quarterly*, 104(2), 1–20.
- 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*, 32(1), 37–58. <https://doi.org/10.1075/target.18157.chm>
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- 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), 71–102. <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 C. Fantinuoli (Ed.), *Interpreting and Technology* (pp. 13–27). Language Science Press.
- EABM. (2021). Survey – Ergonomics for the Artificial Booth Mate (EABM). <https://www.eabm.ugent.be/survey/>
- Fantinuoli, C. (2017). Speech Recognition in the Interpreter Workstation. In J. Esteves-Ferreira, J. Macan, R. Mitkov & O.-M. Stefanov (Eds.), *Proceedings of the 39th Conference Translating and the Computer* (pp. 25–34). ASLING.
- Fantinuoli, C. (2018). Computer-assisted Interpreting: Challenges and Future Perspectives. In G. Corpas Pastor & I. Durán Muñoz (Eds.), *Trends in E-tools and Resources for Translators and Interpreters* (pp. 153–174). Brill.
- Fantinuoli, C. (2023). Towards AI-enhanced Computer-assisted Interpreting. In G. Corpas Pastor & B. Defrancq (Eds.), *Interpreting Technologies: Current and Future Trends* (Vol. 37; pp. 46–72). John Benjamins.
- Frittella, F. M. (2023). *Usability Research for Interpreter-centred Technology*. Language Science Press.
- Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training*. John Benjamins.
- Guo, M., & Han, L. (2024). From Manual to Machine: Evaluating Automated Ear–Voice Span Measurement in Simultaneous Interpreting. *Interpreting*, 26(1), 24–54. <https://doi.org/10.1075/intp.00100.guo>
- Guo, M., Xie, Y., Han, L., Lei, V. L. C., & Li, D. (2025a). Automatic live captioning in Simultaneous Interpreting: A Comparative Analysis of Quality. *Meta*, 70(1), 221–242. <https://doi.org/10.7202/1125409ar>
- Guo, M., Xie, Y., Han, L., Lei, V. L. C., & Li, D. (2025b). Bridging Traditional and AI-assisted Simultaneous Interpreting: Empirical Insights for Curriculum Design. *The Interpreter and Translator Trainer*, 19(3–4), 425–443. <https://doi.org/10.1080/1750399X.2025.2533007>
- Guo, M., Xie, Y., Han, L., Lei, V. L. C., & Li, D. (2025c). From sight to insight: Unravelling student interpreters’ usage and perceptions of live captioning in Chinese–English simultaneous Interpreting. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 160–187. <https://doi.org/10.52034/lans-tts.v24i.834>

- Han, C. (2015a). Lacunae, Myths and Legends about Conference Interpreters: A Diary Study to Explore Conference Interpreting Practice in China. *Perspectives*, 23(3), 440–457. <https://doi.org/10.1080/0907676X.2014.973422>
- Han, C. (2015b). (Para)linguistic Correlates of Perceived Fluency in English-to-Chinese Simultaneous Interpretation. *International Journal of Comparative Literature and Translation Studies*, 3(4), 32–37. <https://doi.org/10.7575/aiac.ijclts.v.3n.4p.32>
- Han, C. (2018). Using Rating Scales to assess Interpretation: Practices, Problems and Prospects. *Interpreting*, 20(1), 59–95. <https://doi.org/10.1075/intp.00003.han>
- Han, C. (2021). Analytic Rubric Scoring versus comparative judgment: A comparison of Two Approaches to Assessing Spoken-language Interpreting. *Meta*, 66(2), 337–361. <https://doi.org/10.7202/1083182ar>
- Han, L., Lu, J., Wen, Z., & Tian, Y. (2023). Momentary Engagement in Simultaneous versus Consecutive Interpreting: Through the Lens of Translanguaging and CDST. *Frontiers in Psychology*, 14, 1180379. <https://doi.org/10.3389/fpsyg.2023.1180379>
- He, Y., Sainath, T. N., Prabhavalkar, R., McGraw, I., Alvarez, R., Zhao, D., Rybach, D., Kannan, A., Wu, Y., Pang, R., Liang, Q., Bhatia, D., Shangguan, Y., Li, B., Pundak, G., Sim, K. C., Bagby, T., Chang, S., Rao, K., & Gruenstein, A. (2019). Streaming End-to-end Speech Recognition for Mobile Devices. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6381–6385). IEEE. <https://doi.org/10.1109/ICASSP.2019.8682336>
- Huang, X., & Zhang, X. (2025). Translating Culture: The Rise and Resonance of Chinese Contemporary Literature in the Portuguese-speaking World. *Humanities and Social Sciences Communications*, 12(1), 168. <https://doi.org/10.1057/s41599-025-04457-z>
- Jiang, L., Jiang, Y., & Han, L. (2024). The Potential of ChatGPT in Translation Evaluation: A Case Study of the Chinese–Portuguese Machine Translation. *Cadernos de Tradução*, 44, e98613. <https://doi.org/10.5007/2175-7968.2024.e98613>
- Jurafsky, D., & Martin, J. H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). Prentice Hall. <https://web.stanford.edu/~jurafsky/slp3>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
- Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.), *HCI and Usability for Education and Work* (pp. 63–76). Springer.
- Lewis, J. R. (1995). IBM computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human–Computer Interaction*, 7(1), 57–78. <https://doi.org/10.1080/10447319509526110>
- Liu, S., Wang, L., & Liu, C.-H. (2018). Chinese–Portuguese Machine Translation: A Study on Building Parallel Corpora from Comparable Texts. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (pp. 1485–1492). European Language Resources Association. <https://aclanthology.org/L18-1236/>

- Luo, W., & Zhang, X. (2025). Lexical Features in the Portuguese Translations of Contemporary Political Discourse: A Corpus-based Analysis. *Forum for Linguistic Studies*, 7(8), 938–948. <https://doi.org/10.30564/fls.v7i8.10785>
- Ma, X., & Cheung, A. K. F. (2020). Language Interference in English-Chinese Simultaneous Interpreting with and without Text. *Babel*, 66(3), 434–456. <https://doi.org/10.1075/babel.00168.che>
- Macao Polytechnic University. (2025, July 4). Macao Polytechnic University Regulations on Research Ethics and Integrity. https://www.mpu.edu.mo/en/dei/guidelines_regulations
- National Health Commission of the People's Republic of China. (2023, February 27). Measures for the ethical review of life science and Medical Research Involving Human Subjects. <https://www.nhc.gov.cn/qjjys/c100016/202302/6b6e447b3edc4338856c9a652a85f44b.shtml>
- Pisani, E., & Fantinuoli, C. (2021). Measuring the Impact of Automatic Speech Recognition on Number Rendition in Simultaneous Interpreting. In C. Wang & B. Zheng (Eds.), *Empirical Studies of Translation and Interpreting* (pp. 181–197). Routledge.
- Pöchhacker, F. (2016). *Introducing Interpreting Studies* (2nd ed.). Routledge.
- Prandi, B. (2023). *Computer-assisted Simultaneous Interpreting: A Cognitive-experimental Study on Terminology*. Language Science Press.
- Saeed, M. A., González, E. R., Korybski, T., Davitti, E., & Braun, S. (2022). Connected yet Distant: An Experimental Study into the Visual Needs of the Interpreter in Remote Simultaneous Interpreting. In M. Kurosu (Ed.), *Human-computer Interaction: User Experience and Behavior* (pp. 214–232). Springer. https://doi.org/10.1007/978-3-031-05412-9_16
- Seeber, K. G. (2017). Multimodal Processing in Simultaneous Interpreting. In J. W. Schwieter & A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 461–475). Wiley-Blackwell. <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>
- Setton, R. (2015). Simultaneous with Text. In F. Pöchhacker (Ed.), *Routledge Encyclopedia of Interpreting Studies* (pp. 385–386). Routledge.
- Shreve, G. M., Lacruz, I., & Angelone, E. (2010). Cognitive Effort, Syntactic Disruption, and Visual Interference in a Sight Translation Task. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 63–84). John Benjamins.
- Silva, I. A. L., Alves, F., Schmaltz, M., Pagano, A., Wong, D., Chao, L., Leal, A. L. V., Quaresma, P., Garcia, C., & Silva, G. E. (2017). Translation, Post-editing and Directionality: A Study of Effort in the Chinese-Portuguese Language Pair. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Translation in Transition: Between Cognition, Computing and Technology* (pp. 107–134). John Benjamins.
- Silva, I. A. L., Schmaltz, M., Alves, F., Pagano, A., Wong, D., Chao, L., Leal, A. L. V., Quaresma, P., & Garcia, C. (2015). Translating and Post-editing in the Chinese-Portuguese Language Pair: Insights from an Exploratory Study of Key Logging and Eye Tracking. *Translation Spaces*, 4(1), 145–169. <https://doi.org/10.1075/ts.4.1.07sil>

- Sun, Y. (2017). Positive Language Transfer of Causative Verbs in Chinese-Portuguese Simultaneous Interpretations. *Letras de Hoje*, 52(3), 322–330. <https://doi.org/10.15448/1984-7726.2017.3.29364>
- Wang, X., Wang, B., & Yuan, L. (2025). The Function of ASR-generated Live Transcription in Simultaneous Interpreting: Trainee Interpreters' Perceptions from Post-task Interviews. *Humanities and Social Sciences Communications*, 12(1), 166. <https://doi.org/10.1057/s41599-025-04492-w>
- Will, M. (2020). Computer-aided Interpreting (CAI) for Conference Interpreters: Concepts, Content and Prospects. *Journal for Communication Studies*, 13(1(25)), 37–71.
- Yang, S., Li, D., & Lei, V. L. C. (2020). The Impact of Source Text Presence on Simultaneous Interpreting Performance in Fast Speeches: Will it Help Trainees or Not? *Babel*, 66(4–5), 588–603. <https://doi.org/10.1075/babel.00189.yan>
- Yu, M., Jiang, X., & Yang, S. (2024). Brazilian sinologist Reis's translation strategy: Comparing his Portuguese translation of Mo Yan's novel *Wa* with Goldblatt's English version. *Theory and Practice in Language Studies*, 14(3), 693–702. <https://doi.org/10.17507/tpls.1403.09>
- Yuan, L., & Wang, B. (2023). Cognitive Processing of the Extra Visual Layer of Live Captioning in Simultaneous Interpreting: Triangulation of Eye-tracking 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*, 22(1), 119–145. <https://doi.org/10.1075/forum.00038.yua>
- 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*, (2), 100131. <https://doi.org/10.56395/ijceti.v2i2.50>

Appendix A. Post-experiment questionnaire

The post-experiment questionnaire comprised 20 main items addressing task difficulty, attention allocation, tool usability, and interface preferences.

Section I: Task difficulty and attention allocation

1. Rate the difficulty of each interpreting task. (7-point scale: 1 = very easy, 7 = very difficult)
2. During SI with text, how did you allocate attention between listening and reading? (Percentage distribution)
3. Was this attention allocation deliberate or unconscious? (Binary choice)
4. During ASR-assisted SI, how did you allocate attention between listening and reading? (Percentage distribution)
5. Was this attention allocation deliberate or unconscious? (Binary choice)
6. How helpful did you find the textual support for each SI mode? (7-point scale: 1 = not helpful, 7 = very helpful)



Section II: Task comparison

7. How similar is ASR-assisted SI to other interpreting modes? (7-point scale: 1 = not similar, 7 = very similar)
8. How similar is SI with text to other interpreting modes? (7-point scale: 1 = not similar, 7 = very similar)

Section III: ASR usability (adapted from SUS; Brooke, 1996)

9. Rate your agreement with 10 statements regarding ASR ease of use, learnability, and confidence. (5-point scale: 1 = strongly disagree, 5 = strongly agree)

Section IV: ASR quality and interface (adapted from CSUQ; Lewis, 1995)

10. Rate your agreement with 11 statements regarding ASR accuracy, information clarity, interface layout, and overall satisfaction. (7-point scale: 1 = strongly disagree; 7 = strongly agree)

Section V: ASR latency and interference

11. Rate six aspects of ASR performance including perceived latency, disruption, helpfulness, and audio-text alignment. (7-point scale: 1 = not at all, 7 = very much)

Section VI: Subjective experience (adapted from UEQ; Laugwitz et al., 2008)

12. Rate your subjective experience with the ASR tool on eight semantic differential pairs (e.g., obstructive–supportive, complicated–simple, inefficient–efficient). (7-point scale: –3 to +3)

Section VII: Technology acceptance (adapted from TAM; Davis, 1989)

13. Rate your agreement with eight statements regarding perceived usefulness and ease of use of digital interpreting tools. (5-point scale: 1 = strongly disagree, 5 = strongly agree)

Section VIII: Interface preferences

14. Preferred device for ASR use during SI. (Multiple choice)
15. Preferred ASR interface format for on-site interpreting. (Multiple choice)
16. Preferred ASR interface format for online interpreting. (Multiple choice)
17. Preferred display format for standalone ASR window. (Multiple choice)
18. Preferred display format for subtitle-style ASR. (Multiple choice)

Section IX: Open-ended questions

19. Are there other computer-assisted tools suitable for use during SI? (Free response)
20. Additional comments or suggestions. (Free response)

Appendix B. Retrospective Interview Protocol

Retrospective interviews were conducted after each interpreting task and at the conclusion of the experimental session. The protocol comprised three parts: post-task ratings administered



after each condition, follow-up questions for tasks involving visual-verbal support, and summative questions addressing overall experience.

Part I: Questions for tasks with visual-verbal support

1. How often did you refer to the visual-verbal input (transcription/text) during this task?
2. Were there moments when the visual-verbal input was especially helpful or distracting? Please describe.
3. How did the presence of visual-verbal information affect your overall strategy, accuracy, and confidence?
4. Did the visual-verbal input lead to any adjustments in your interpreting strategies, especially when confronted with ambiguous or complex segments?
5. Were there any discrepancies or conflicts between the auditory input and the visual-verbal input? How did you manage these situations?
6. To what extent did the visual-verbal input influence your confidence and ease during the task?
7. In your experience, did the visual-verbal input predominantly act as an aid or a complication? Please explain.
8. Were there instances where you delayed your interpretation, waiting for the visual input to clarify or confirm the message?
9. What recommendations or improvements would you suggest for the transcription tool? (ASR condition only)

Part II: Summative questions (administered after all tasks)

10. Looking back, which task imposed the highest cognitive load, and why?
11. As you switched between different interpreting conditions, did you notice any shifts in your cognitive strategy or focus?
12. From a training perspective, do you believe that experience in other interpreting modes can benefit performance in ASR-assisted SI?
13. Based on your experience, what do you foresee as potential challenges or developments in computer-assisted simultaneous interpreting?
14. Are there any additional insights, feedback, or suggestions you would like to offer?

Editorial notes

Authorship contribution

Conceptualization: M. Guo, L. Han, M. T. Anacleto

Data collection: M. Guo

Data analysis: M. Guo

Results and discussion: M. Guo, L. Han, M. T. Anacleto

Writing – review and editing: M. Guo, L. Han, M. T. Anacleto

Research dataset

The data analyzed in this study were collected as part of a doctoral research project conducted by the first author at Macao Polytechnic University.



Funding

This work was supported by the Fundamental Research Funds for the Central Universities in the University of International Business and Economics under Grant [24QD32].

Image copyright

Original experimental photographs and screenshots are the authors' own or were used with the consent of the people depicted, where applicable. All non-original visual materials are credited in the relevant figure notes as adapted or reconstructed from the cited sources. Permission was obtained where required.

Approval by ethics committee

Not applicable. This minimal-risk behavioral study followed institutional principles on research integrity and participant protection, now formalized in the *Macao Polytechnic University Regulations on Research Ethics and Integrity*, together with institutional guidance for investigations involving human participants. Written informed consent was obtained from all participants; confidentiality and the right to withdraw without penalty were guaranteed; no sensitive personal data were collected, and no deception was used. National ethical principles on informed consent and reasonable participant compensation were also observed (NHC, 2023, Arts. 17(4) and 33).

Conflicts of interest

Not applicable.

Data availability statement

The data from this research, which are not included in this work, may be made available by the author(s) upon request.

License

The authors grant *Cadernos de Tradução* exclusive rights for first publication, while simultaneously licensing the work under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#) License. This license enables third parties to remix, adapt, and create from the published work, while giving proper credit to the authors and acknowledging the initial publication in this journal. Authors are permitted to enter into additional agreements separately for the non-exclusive distribution of the published version of the work in this journal. This may include publishing it in an institutional repository, on a personal website, on academic social networks, publishing a translation, or republishing the work as a book chapter, all with due recognition of authorship and first publication in this journal.

Publisher

Cadernos de Tradução is a publication of the Graduate Program in Translation Studies at the Federal University of Santa Catarina. The journal *Cadernos de Tradução* is hosted by the [Portal de Periódicos UFSC](#). The ideas expressed in this paper are the responsibility of its authors and do not necessarily represent the views of the editors or the university.

Section editor

Willian Moura

Style editors

Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

Article history

Received: 11-08-2025

Approved: 26-01-2026

Revised: 20-02-2026

Published: 09-2026

