Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load

Zhibin Yu
Zhejiang Gongshang University
Hangzhou, Zhejiang Province, China
zhibin88yu@gmail.com
https://orcid.org/0009-0003-5618-497X

Zhangminzi Shao (corresponding author)*
Zhejiang Gongshang University
Hangzhou, Zhejiang Province, China
zhangminzi.shao@hotmail.com
https://orcid.org/0000-0003-3339-8062

Abstract

This study investigated the impact of AI-powered automatic speech recognition (ASR) technology on interpreting performance and cognitive load during consecutive interpreting (CI), particularly in the case of source speeches that feature varying accents. Multiple performance metrics – fidelity, fluency, target language (TL) quality and overall quality – were assessed. Fluency was measured by means of delivery rate, frequency and mean length of silent pauses, and the overall occurrence of disfluencies. Cognitive load was evaluated using subjective self-ratings and an objective measure of fundamental frequency (*F0*). Twenty-four advanced student interpreters conducted four CI tasks each – two with an unfamiliar accent and two without one – both with and without ASR assistance. The results indicate that the impact of ASR on CI is multifaceted. Whereas ASR improved the interpreting fidelity, it reduced the delivery rate and increased the frequency of silent pauses. This complexity is more pronounced when the source speech features an unfamiliar accent, which could lead to interpreters' over-reliance on ASR and as a result compromise the TL quality. No significant effect was observed on the overall quality of the interpreting. Notably, ASR did not affect the interpreters' cognitive load across all the phases of CI, regardless of the presence of unfamiliar accents in the source speeches.

Keywords: automatic speech recognition, ASR, consecutive interpreting, accent, interpreting performance, cognitive load

1. Introduction

Recent advances in deep learning and neural networks have significantly improved automatic speech recognition (ASR) technology, achieving impressively low word-error rates (WERs) and reduced latency (Wang et al., 2020), even for heavily accented speech (Ghangam et al., 2021; Hacking et al., 2023; Li, 2022; Romero et al., 2024). These improvements have extended the application of ASR beyond the preparation stage to help interpreters during the interpreting process. ASR now provides real-time transcriptions of source speech and has been widely adopted in various interpreting settings, including conferences, courts, and liaison interpreting (Fantinuoli, 2017; Gerber et al., 2020; Pöchhacker, 2016; Will, 2020; Yuan & Wang, 2023). In these ways, it demonstrates its "considerable potential for changing the way interpreting is practiced" (Pöchhacker, 2016, p. 188).

Given the extended application of ASR in interpreting practice, an increasing number of studies have been conducted on the ways in which ASR may influence interpreting processing and products. Some research investigated the ways ASR-based computer-assisted interpreting (CAI) tools contribute to resolving specific "problem triggers" during simultaneous interpreting (SI). For instance, it has been reported that InterpretBank improved the accuracy of interpreting both numbers (Defrancq & Fantinuoli, 2021; Pisani & Fantinuoli, 2021) and terminology (Prandi, 2018, 2023). Other studies probed into ASR's influences on simultaneous interpreters' cognitive processing, exploring how live captioning may affect attention allocation and cognitive load (Prandi, 2023; Yuan & Wang, 2023). Despite these insightful findings reported in the literature, most of them were obtained from research on SI, leaving the impacts of ASR support in consecutive interpreting (CI) underexplored.

Three critical issues regarding the application of ASR technology in CI have yet to be investigated. First, its overall impact on CI quality is still unclear. The few empirical studies that have tested the effectiveness of ASR-generated transcripts and their machine translations reported higher accuracy but lower fluency (e.g., Ünlü, 2023). However, to date, no study has provided a systematic quality profile – one that simultaneously rates fidelity, target-language (TL) quality, delivery fluency, and overall performance – under ASR assistance.

Second, ASR's impact on consecutive interpreters' cognitive load needs to be explored further. Only Chen and Kruger (2023) have empirically examined the effect of ASR on cognitive load in a novel mode of computer-assisted CI (CACI) that integrates both speech recognition and machine translation (MT), replacing note-taking with simultaneous respeaking. They reported a reduced cognitive load when student interpreters worked from L1 to L2. However, this experimental setup differs substantially from conventional CI practices, which may limit the generalizability of their findings.

Third, research has rarely examined whether ASR can mitigate the cognitive challenges posed by unfamiliar accents. An unfamiliar accent is likely to reduce the intelligibility of source speech, leading to a possible increase in interpreters' cognitive load (e.g., Kurz, 2008; McAllister, 2000) and possibly the impairment of the accuracy of their output (Lin et al., 2013). ASR may be a

promising technique to deal with this challenge, as in such cases its live captioning may aid comprehension during listening and analysis; moreover, its transcripts could support sight translation during production (Fantinuoli, 2017). But this hypothesis has not yet been tested. The present study aims to respond to these matters by investigating the potential benefits of AI-powered ASR technology to CI, focusing as it does on its effects when handling speeches with varying accents.

2. Accent interference in interpreting

Accents are distinct ways in which a language is pronounced, whether by native or by non-native speakers (Levis & Zhou, 2018). According to Lippi-Green (2011), they are "loose bundles of prosodic and segmental features distributed over geographic and/or social space" (p. 42). Accents can be classified as either local and/or regional (i.e., L1 accent) or foreign (i.e., L2 accent) (Pérez-Ramón et al., 2022).

The spread of global English has led to the increasing presence of non-native English speech in conference settings (Albl-Mikasa, 2022; Chang & Wu, 2014). According to a survey by Albl-Mikasa (2010, p. 142), 71% of the respondents, who were professional interpreters who mostly speak German, French, Dutch, or Italian as their native language, reported encountering foreign accents "very frequently" in Europe. In China, where interpreters are also faced with a wide variety of English accents (Cheung & Li, 2022), the prevalence of L2 English speakers has become normal at conferences (Chang & Wu, 2014). Moreover, research has also indicated that certain L2 English speakers tend to speak with their L1-influenced accent in order to express certain desired identities (Sung, 2016).

Research from the field of English as a second language (ESL) or English as a foreign language (EFL) has shown that whereas foreign-accented speech may not necessarily result in low intelligibility (i.e., being difficult or impossible to understand), it often requires more processing time and places a higher cognitive load on listeners (Munro & Derwing, 1995). If the task itself already poses high cognitive demands on the listener, challenging accents would have a greater impact on intelligibility (Jensen & Thøgersen, 2017). In addition, the challenge posed by L2 speakers depends on the listener's familiarity with that accent (Gass & Varonis, 1984). If the listener is used to the accent, comprehension is easier, whereas unfamiliar accents can significantly increase the cognitive effort required to gain an understanding of the speech.

In the field of interpreting, accents are known to be one of the challenging factors that may have a negative impact on interpreting performance, especially in SI (Albl-Mikasa et al., 2017; Basel, 2003; Gile, 2009). Non-native or unfamiliar accents take longer for interpreters to understand and therefore increase the interpreters' cognitive load. When more cognitive resources are consumed by comprehension, fewer resources are available for other interpreting subtasks, which leads to a greater likelihood of errors and omissions (Cheung & Li, 2022). Findings from previous empirical studies have lent support to this proposition. Lin et al. (2013) found that non-native accents in source speeches reduced the accuracy of rendition in SI. Studies by McAllister (2000), Sabatini (2000) and Kurz (2008) have shown that speakers' non-native accents added to student interpreters' cognitive load.

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

However, it is important to note that not all non-native accents are equally challenging. Interpreters are likely to understand ELF speakers from a common L1 background more easily, whereas they would find those with significantly different language backgrounds difficult to interpret (Chang & Wu, 2014; Katikos, 2015). In addition, the "shared languages benefit" (Albl-Mikasa, 2013, p. 105) – that is, the interpreter's familiarity with the speaker's L1 – would facilitate comprehension of the source speech and therefore streamline the interpreting process (Kurz & Basel, 2009).

For interpreters from countries or regions with relatively homogeneous linguistic environments where English is not the native language, non-native English accents can pose significant challenges. Cheung and Li (2022) have highlighted this issue for Chinese interpreters, noting that their English training often focused on native-English norms with a heavy emphasis on standard pronunciation and limited exposure to different types of English accent.

3. ASR as an interpreter assister

Automatic Speech Recognition (ASR) is a computer-based method of converting audio recorded speech to written text (Gutz et al., 2022). In recent years, significant advancements in ASR technology have led to its widespread application across various fields: for instance, modern vehicles (Park & Na, 2020), home assistant devices (Zhu et al., 2024), speech usability for the deaf and hard-of-hearing (Glasser, 2019), and computer games (Mustaquim, 2013).

Early efforts to integrate ASR into the interpreting workflow aimed at automating the querying system of CAI tools as a documentation aid (Fantinuoli, 2017; Gerber et al., 2020). However, as the technology has evolved, its applications have extended beyond the preparation phase and are now actively used during the interpreting process itself (Fantinuoli, 2017).

In traditional SI settings, interpreters may rely heavily on their booth mates to deal with challenging information elements such as numbers, terms and proper names, and to communicate with clients and technicians (Seresi & Láncos, 2022). With the integration of ASR-based software – for instance, InterpretBank (Defrancq & Fantinuoli, 2021; Pisani & Fantinuoli, 2021), Kudo Interpreter Assist (Fantinuoli et al., 2022), Zoom ASR subtitles (Yuan & Wang, 2023), iFlytek (Sun et al., 2021) and SmarTerp (Frittella & Rodríguez, 2022) – this technology is expected to play the role of an artificial booth mate (Fantinuoli, 2023). In this way, it would facilitate interpreting tasks either by providing automatic transcription of the source text or specific information (Defrancq & Fantinuoli, 2021; Li & Chimel, 2024) or by combining MT technology to offer both the source text and its machine-translated output of the transcription (Su & Li, 2023; Sun et al., 2021). Empirical efforts have been made to evaluate the effectiveness of ASR-based technological support either in simulated environments or with fully developed systems. Numerical and term accuracy in output have often been the research focus (Defrancq & Fantinuoli, 2021; Desmet et al., 2018; Fantinuoli & Montecchio, 2023; Frittella & Rodríguez, 2022; Pisani & Fantinuoli, 2021; Tammasrisawat & Rangponsumrit, 2023; Yuan & Wang, 2023). For instance, Desmet et al. (2018) simulated a scenario in which ASR displayed source-language (SL) numbers directly in front of interpreters during SI and found that the accuracy of the interpreted numbers increased significantly from 56.5% to 86.5%. Similarly, Defrancq and Fantinuoli (2021), using the InterpretBank ASR as the CAI tool, found through empirical

research that ASR assistance improved the overall quality for nearly all types of numbers during SI. Moreover, Prandi (2023) reported increased term accuracy in the ASR-assisted SI condition. However, some studies have also argued that while ASR assists interpreters during SI, it may reduce the fluency of their output (Cheung & Li, 2022; Tammasrisawat & Rangponsumrit, 2023).

A few studies have explored the impact of ASR on the cognitive processing of interpreters in SI. It is generally assumed that the use of ASR–CAI tools may require supplementary cognitive resources because interpreters have to process an additional visual input (Prandi, 2018; Will, 2015). In contrast, Defrancq and Fantinuoli (2021) and Cheung and Li (2022) have described the ASR tool as a "safety net", suggesting that it may have a beneficial psychological effect. Furthermore, based on the analysis of theta power in EEG signals, Li and Chmiel (2024) reported a reduction in cognitive load when interpreting with ASR subtitles and argued that the cognitive cost of processing the additional information channel is offset by the cognitive gain achieved through visual aids.

Compared to SI, there have been fewer discussions and empirical explorations of the application of ASR in CI. Most existing studies have incorporated ASR into the CI workflow, using it to provide real-time transcripts of the source speech – either as a standalone aid or as input for subsequent MT – in order to support interpreters during a CI task (Goldsmith & Blouin, 2021; Ünlü, 2023; Wang & Wang, 2019). In these studies, interpreters were provided with either (a) ASR-generated transcripts of the source speech (Goldsmith & Blouin, 2021) or (b) MT outputs of those transcripts (Wang & Wang, 2019), or (c) both the ASR transcripts and their corresponding MT outputs (Ünlü, 2023).

Some ASR-based software also changes the way interpreters take notes in CI, shifting away from the traditional pen and paper and instead using a stylus on a tablet. For example, Cymo Note's CI feature allows interpreters to divide the screen into two parts: with the transcription on the left and a blank space for note-taking on the right. In this setup, interpreters can use the drawing mode to write notes anywhere on the page. Sight-Terp offers two automatically generated reference texts: the source text on the left and the target text on the right. These are accompanied by a digital notepad that supports stylus-based note-taking. An exception to this trend is the CACI model proposed by Chen and Kruger (2023), which integrates speech recognition (SR) and MT in a fundamentally different way. In this model, interpreters listen to a source speech and repeat it into the SR system, which generates an SR transcript. This transcript is then processed by the MT system and the interpreters use both the SR and the MT texts as references for interpreting.

Despite these innovations, only two studies have examined how the application of MT or ASR technology would affect CI products, both of them reporting gains in accuracy but also new sources of distraction or disfluency. Wang and Wang (2019) assessed ten student interpreters' English-to-Chinese CI performance, comparing the results under two conditions: CI with MT assistance and the traditional CI mode without any assistance. Five student interpreters performed better in accuracy with MT support, although their fluency did not show a clear trend. The post-task interviews revealed that whereas most interpreters found the MT reference helpful, eight out of ten viewed it as a distraction and nine out of ten struggled to locate the relevant information owing to the full-paragraph display format.

The second study was that by Ünlü (2023), who designed Sight-Terp, a web-based ASR-enhanced CAI tool, and conducted empirical research on its effectiveness. The tool initiates continuous speech recognition to transcribe the source speech input and automatically generates machine-translated outputs of speech segments, creating two automatically generated reference texts. The results show a significant improvement in the accuracy of the content but an increase in disfluency markers, which could be attributed to the cognitive load caused by processing additional information from the tool.

Apart from these revealing findings, several aspects remain unexplored. First, to date, there has been only limited exploration of simpler ASR-assisted CI methods that involve fewer electronic tools, such as providing interpreters only with real-time SL transcripts while allowing them to continue using traditional pen-and-paper note-taking. Given the increasing availability of modern ASR tools and the fact that digital pens are both more expensive and require more adaptation for note-taking than traditional pens (Orlando, 2015), this alternative approach deserves further investigation.

Second, the impact of ASR assistance on interpreters' cognitive processing remains to be examined empirically. In theory, ASR may reduce the effort required for subtasks such as listening and analysis, note-taking, and recalling from short-term memory. This is because the transcripts generated may facilitate comprehension and note-taking and, in the production phase, it may be used for sight translation (Fantinuoli, 2023; Gerber et al., 2020; Goldsmith & Blouin, 2021). However, the visual input derived from reading the transcript is also likely to create new cognitive demands, offsetting or even outweighing the gains. To date, no research has examined this matter empirically. Detailed analyses of interpreters' cognitive load at distinct stages of the CI process – namely, while listening and note-taking, and during target-speech production (Gile, 2009) – or for specific subtasks, are still lacking. Whether ASR can really reduce interpreters' cognitive load remains to be empirically validated.

Third, for language pairs such as English and Chinese that differ significantly in structure, interpreters often need to restructure information to align it with the target audience in order to ensure effective communication. Whereas ASR enables interpreters to have the source speech transcription before their eyes, this also introduces the risk of SL interference (Agrifoglio, 2004). Interpreters may unintentionally rely too heavily on the structure or phrasing of the SL, rather than adapting it to the TL. Little is known about the influence of ASR technology on TL quality and even less about its effect on overall interpreting quality.

Moreover, no research has investigated ASR-assisted CI in the context of accented source speeches. While advanced ASR systems have improved their handling of accented speech, it remains unclear whether ASR support can effectively mitigate the cognitive challenges that interpreters face when working with unfamiliar or heavily accented speakers.

4. The present study

The present study aimed to respond to the following two research questions:

- **RQ** 1: Does ASR support improve interpreters' performance during CI compared to CI without ASR support? If so, does this effect differ between accented and non-accented conditions?
- **RQ** 2: Does the use of ASR assistance in CI reduce interpreters' cognitive load compared to CI without ASR support? If so, does this effect differ between accented and non-accented conditions?

A group of advanced interpreting students interpreted four speeches consecutively from their B language into their A language – two featuring a challenging accent and two without. Each interpreter had to work with two speeches with ASR assistance and two without, each speech having been assigned to only one condition per interpreter. The participants' CI performances and cognitive loads were to be compared between the two ASR conditions and also between the two accent conditions in order to respond to the two research questions. Their CI performance was assessed using four criteria:

- fidelity
- TL quality
- fluency and
- overall quality (for details, see section 5.6).

Cognitive load was measured using both a subjective questionnaire and an objective indicator of fundamental frequency (*F0*) – specifically the mean *F0* (for details, see section 5.7).

Regarding their CI performance and drawing on previous findings (Ünlü, 2023; Wang & Wang, 2019), we hypothesized that ASR assistance would increase fidelity but reduce fluency. We also expected it to reduce the TL quality while enhancing overall quality. Finally, we predicted that the impact of ASR would be more pronounced when the source speeches were accented. As regards cognitive load, we hypothesized that with ASR assistance the interpreters would report reduced pressure on both listening and note-taking, in addition to a reduced cognitive load during TL production. We likewise expected a lower mean *F0* in the ASR condition. Finally, we predicted that all of these effects would be more pronounced when the source speeches were accented.

5. Method

5.1 Participants

The study involved 24 student interpreters from a key provincial university in eastern China, all of them graduate students specializing in interpreting. They were paid after the experiment. At the time of the experiment, they had completed three semesters of conventional CI training and one semester of SI training. All of the participants are native Chinese speakers and English is their B language. Each participant had obtained either a Level II or a Level III *Certificate of China Accreditation Test for Translators and Interpreters* (CATTI). All of the participants had

experience in learning at least one foreign language besides English. Among them, 17 had studied French, four had studied German, ten had studied Japanese, and two had studied Spanish. The 24 participants included five males and 19 females, their ages ranging from 21 to 31 years (M = 24.75, SD = 2.01).

The participants were informed during their recruitment that they would undergo five ASR-assisted CI training sessions prior to the experiment. Each session lasted two hours and the participants were allowed to withdraw at any time during the training. All of the participants provided written consent and then completed the training sessions before participating in the experiment.

5.2 Training

Five ASR-assisted CI training sessions were provided, each lasting two hours and separated by a one-week interval, covering a total duration of ten weeks. The training took place in a standard interpreting classroom. The initial session featured a lecture on the application of ASR in CI workflows. Various ASR software options were evaluated based on criteria that included latency, WER, ease of use, and cost-effectiveness. Otter.ai was selected as the ASR tool for the following English–Chinese ASR-assisted CI session. The subsequent four sessions focused on ASR-assisted CI across diverse scenarios and topics. The training materials covered fields such as education, economics, politics, and technology, and varied in difficulty according to multiple dimensions. After each session, the participants received additional materials for practice and feedback on their assignments. In addition, after each session, interviews were conducted with three randomly selected participants to gather insights into their ASR experience and their perceptions of the interpreting process. This pre-experimental exposure to ASR-assisted CI, systematically implemented over an extended period and embedded in authentic training contexts, was intended to mitigate substantially the potential influence of the "novelty effect" on the participants' performance during the experimental phase.

5.3 Materials

Four English speeches were prepared for the experiment. Speech 1 and speech 3 were delivered by a Brazilian Ph.D. in Neuro-management with a Portuguese accent (PAE), while speech 2 and speech 4 were delivered in General American English (GAE) by a U.S. native English speaker with a Ph.D. in Psychology. A preliminary survey of the participants' language backgrounds confirmed that none of them had studied or been exposed to Portuguese.

All four speeches were similar in terms of topic, structure, delivery rate, and duration, which was approximately three minutes each. Prior to recording their speeches, the two speakers were asked to familiarize themselves with the speech texts. For the purposes of the recording, they were asked to deliver the speeches as if addressing a live audience, aiming for a natural and engaging delivery style and avoiding mechanical word-for-word reading. Table 1 below provides the details of the properties of the four speeches.

Three interpreter trainers rated the speeches on a 7-point Likert scale (1 = very easy to 7 = very difficult), evaluating their lexical and syntactic difficulty, information density, logic, delivery speed, and knowledge difficulty. Their ratings indicated that the difficulty of the four speeches

was consistent across these dimensions except for accent. The assessors rated speeches 1 and 3 as "difficult" in terms of accent, with an average score of 6.33, and speeches 2 and 4 as "easy" in terms of accent, with an average score of 1.67. Neither author participated in the level-of-difficulty assessment.

Table 1

Properties of source speeches

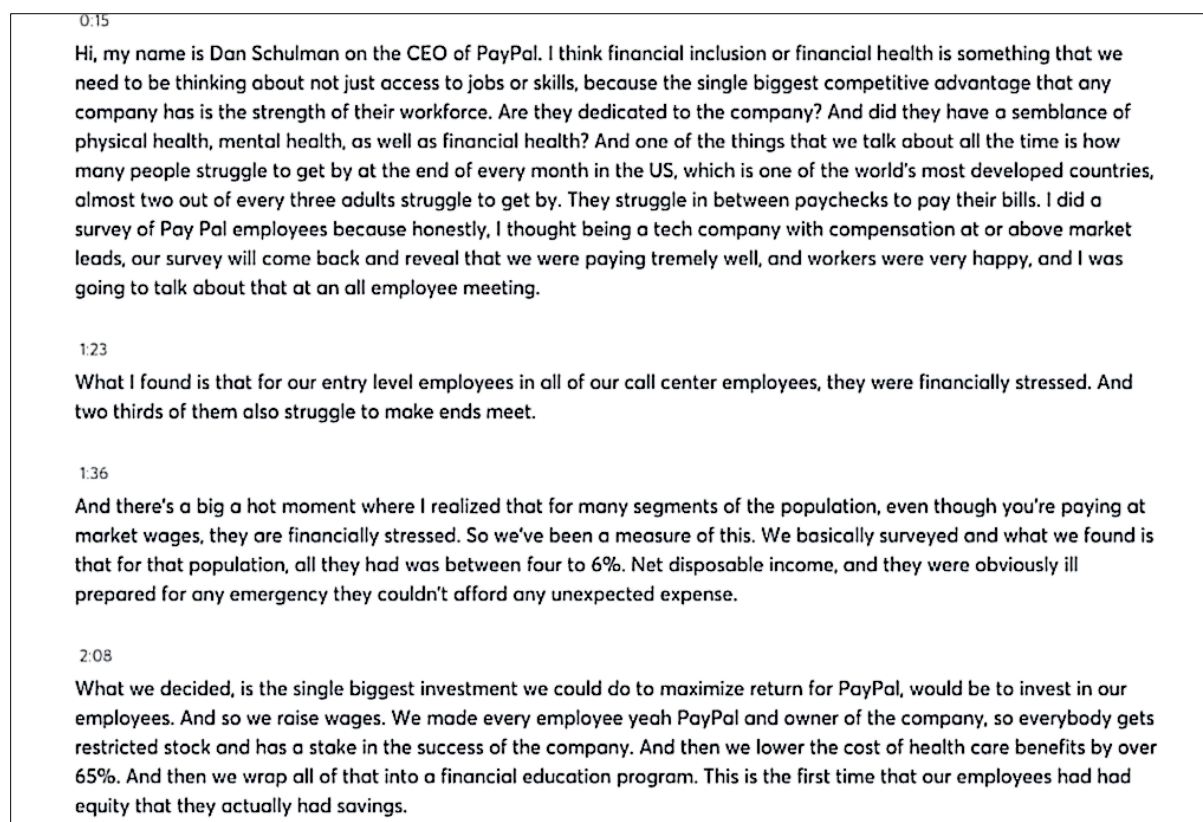| Speech number	Speech topic	Accent	Length (words)	Duration (mins)	Speech rate (wpm)
1	Four-day workweek	PAE	347	02:31	137.9
2	Wage raise	GAE	343	02:35	132.8
3	Work with AI for employees	PAE	374	02:39	141.1
4	AI-Era strategies for employers	GAE	358	02:28	145.1

5.4 ASR assistance

Otter.ai was chosen as the ASR technology for this study. To ensure a consistent ASR experience despite potential network instability and variability in real-time transcription quality, pre-recorded ASR-assisted videos were used for all four speeches. This approach ensured that the participants received identical ASR outputs for the same source texts. The transcription accuracy of Otter.ai for the four speeches was satisfactory, with the WERs as follows: 5.76% for speech 1, 2.09% for speech 2, 2.95% for speech 3, and 1.64% for speech 4. Although Speech 1 exhibited a higher WER, a qualitative analysis showed that 12 of the 19 errors involved function words – such as articles, singular and/or plural forms of third-person nouns, and prepositions – which had a minimal impact on intelligibility or the accurate conveyance of meaning. Recognition latency remained below one second throughout. None of the participants reported any problems with the ASR output after the experiment.

During the CI tasks, the ASR output scrolled across the screen in real-time. Because each speech lasted approximately 2.5 minutes, the full transcription was presented towards the end of the source speech, so the interpreters did not need to scroll manually. Figure 1 displays the complete ASR transcript for speech 2 as it appeared to the interpreters when they produced the TL output.

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

Figure 1

Screen display of ASR for speech 2

> 0:15
>
> Hi, my name is Dan Schulman on the CEO of PayPal. I think financial inclusion or financial health is something that we need to be thinking about not just access to jobs or skills, because the single biggest competitive advantage that any company has is the strength of their workforce. Are they dedicated to the company? And did they have a semblance of physical health, mental health, as well as financial health? And one of the things that we talk about all the time is how many people struggle to get by at the end of every month in the US, which is one of the world's most developed countries, almost two out of every three adults struggle to get by. They struggle in between paychecks to pay their bills. I did a survey of Pay Pal employees because honestly, I thought being a tech company with compensation at or above market leads, our survey will come back and reveal that we were paying tremely well, and workers were very happy, and I was going to talk about that at an all employee meeting.
>
> 1:23
>
> What I found is that for our entry level employees in all of our call center employees, they were financially stressed. And two thirds of them also struggle to make ends meet.
>
> 1:36
>
> And there's a big a hot moment where I realized that for many segments of the population, even though you're paying at market wages, they are financially stressed. So we've been a measure of this. We basically surveyed and what we found is that for that population, all they had was between four to 6%. Net disposable income, and they were obviously ill prepared for any emergency they couldn't afford any unexpected expense.
>
> 2:08
>
> What we decided, is the single biggest investment we could do to maximize return for PayPal, would be to invest in our employees. And so we raise wages. We made every employee yeah PayPal and owner of the company, so everybody gets restricted stock and has a stake in the success of the company. And then we lower the cost of health care benefits by over 65%. And then we wrap all of that into a financial education program. This is the first time that our employees had had equity that they actually had savings.

5.5 Procedure

The English to Chinese CI experiment took place in a standard interpreting classroom compliant with the ISO standard. The participants were not informed in advance of the topics or the inclusion of accents in the experiment. The 24 student interpreters were randomly divided into four groups, with a Latin Square design applied to assign the speeches to each experimental condition. Each group completed four CI tasks from English to Chinese, with two speeches delivered in GAE and two in PAE. Within each accent condition, one speech was presented with ASR assistance and the other without it. This resulted in four experimental conditions: (1) GAE with ASR assistance; (2) GAE without ASR assistance; (3) PAE with ASR assistance; and (4) PAE without ASR assistance (see Table 2).

Table 2

Assignment and conditions for CI tasks

	Task 1	Task 2	Task 3	Task 4
Group 1	Speech 1 (without ASR)	Speech 3 (with ASR)	Speech 4 (with ASR)	Speech 2 (without ASR)
Group 2	Speech 1 (with ASR)	Speech 4 (without ASR)	Speech 3 (without ASR)	Speech 2 (with ASR)
Group 3	Speech 2 (without ASR)	Speech 4 (without ASR)	Speech3 (with ASR)	Speech 1 (with ASR)
Group 4	Speech 2 (with ASR)	Speech 3 (without ASR)	Speech 4 (without ASR)	Speech 1 (with ASR)

Owing to the participants' varying availability, nine interpreting sessions were arranged. Each session involved interpreters from the same group, with a maximum of six participants per session to ensure sufficient space and minimal distractions. The process was identical across all the sessions, with the second author of this article hosting the entire experiment.

Prior to each session, the *Interpreters' Cognitive Load Scale* was distributed to each participant. The participants were briefed on the four CI tasks – two with ASR assistance and two without. Before each task, the participants were given three minutes to prepare. During this time, a PowerPoint presentation displayed on the screen the topic and the key terms of the interpreting task, an indication of ASR availability, and a 20-second audio introduction by the speaker to familiarize the participants with their accent.

During the interpreting sessions, the participants' interactions with the ASR system were observed and recorded. After each task, the participants completed a cognitive load scale specific to that task. For each task, a five-minute break followed the completion of the questionnaire. The total duration of the experiment was approximately 70 minutes.

5.6 Measuring CI performance

As mentioned in section 4, the evaluation of the participants' interpreting performance encompassed several criteria, including fidelity, delivery fluency, TL quality, and overall quality. Fidelity refers to the informational correspondence between what a speaker delivers and what an interpreter renders (Gile, 1995; Han et al., 2021). TL quality is concerned with the grammatical and idiomatic appropriateness and the rendition (Gile, 1995). Fluency reflects the smoothness and continuity of delivery, and overall quality provides a holistic judgement of the interpreter's performance. Fidelity, TL quality and overall quality were evaluated subjectively; fluency was assessed based on several indicators, including the delivery rate, the frequency and the mean duration of silent pauses, plus the occurrence of any disfluencies (Hamidi & Pöchhacker, 2007; Han et al. 2020). Disfluencies were categorized as filled pauses, false starts, repetitions, and slips of the tongue (Tissi, 2000).

Three professional trainers of interpreting were invited to provide a holistic assessment of the fidelity, TL quality and overall quality of the interpretations. They were provided with clear

definitions and detailed scoring guidelines for each evaluation criterion to ensure consistent and standardized rating. Each rater received the source texts along with the audio recordings of all the participants' interpretations. A five-point rating scale was applied, with each point corresponding to a defined score range for a more nuanced assessment: 9–10 for "Excellent" (5), 7–8 for "Good" (4), 5–6 for "Average" (3), 3–4 for "Poor" (2), and 1–2 for "Weak" (1). All three judges were seasoned interpreters with Mandarin as their A language and English as their B language. Neither of the authors participated in the assessment of CI performance. All of the judges rated the CI interpretations independently and the results exhibited a relatively high level of reliability, with an intra-class correlation coefficient (ICC) of 0.78. None of the judges was informed about the conditions and purpose of this experiment. Fluency was assessed using specific indicators, including the delivery rate, the frequency and the mean duration of silent pauses, and the occurrence of disfluencies. Audacity was employed to measure the frequency and mean duration of silent pauses, with a threshold of 0.25 seconds having been set for identifying silent pauses (Towell et al., 1996).

5.7 Measuring cognitive load

Cognitive load was measured using subjective questionnaires which targeted the two main stages of CI: (1) listening and note-taking, and (2) production. The first stage was broken down into listening and analysis, and note-taking. An *Interpreters' Cognitive Load Scale*, adapted from Paas (1992), was implemented. In using this nine-point Likert scale ranging from 1 (very very low) to 9 (very very high), the participants were instructed to self-rate their cognitive load after interpreting each speech, based on three distinct sources respectively: SL listening comprehension, note-taking, and target speech production.

In addition, the mean *F0* was used as a quantitative indicator of cognitive load during the interpreters' production phase. The validity of this measure is supported by previous research on cognitively demanding tasks, such as flight-simulation studies (Boyer et al., 2018; Huttunen et al., 2011; Johannes et al., 2007). The *F0* data for each participant's interpretation were obtained using Praat (Boersma & Weenink, 2023), with a default sampling rate of 100 frames per second. The default pitch range was set at the standard 75–500 Hz for the female participants and 50–400 Hz for the male participants. Mean *F0* data were manually collected through the Pitch Info windows.

5.8 Data analysis

One participant's data were not accepted for analysis due to the failure to record their interpreting performance. Repeated-measures ANOVAs were carried out on different indicators of interpreting performance and cognitive load separately, with accent and ASR assistance as within-subject factors. Greenhouse-Geisser correction was adopted to obtain the adjusted critical values when an assumption of sphericity was violated. When significant interaction between the two factors was found to exist, simple effect analysis would be conducted with Bonferroni correction applied.

6. Results

6.1 Consecutive interpreting performance

Descriptive statistics (means and SDs) for the participants' CI performances are reported in Table 3, and the results of repeated-measures ANOVAs are summarized in Table 4.

6.1.1 Fidelity

A significant main effect of ASR assistance on fidelity was observed [F (1, 22) = 7.18, p = .014, η_p^2 = .246], which indicates a notable enhancement of the completeness and accuracy of the information in the participants' CI products due to ASR support. Furthermore, the absence of a significant interaction effect suggests that this contributory impact of ASR assistance was consistent, regardless of whether the SL speech featured an unfamiliar accent.

6.1.2 Target language quality

No main effect was found on TL quality. However, a significant interaction effect between accent and ASR assistance was observed [F (1, 22) = 4.68, p = .042, η_p^2 = .175]. Subsequent simple effect analysis showed that TL quality was adversely affected when ASR assistance was provided [F (1, 22) = 5.06, p = .035], but this interference effect was pronounced only in the accented condition.

6.1.3 Delivery fluency

The results of repeated-ANOVAs on the four indicators of delivery fluency all revealed a significant or marginally significant main effect of ASR assistance. To be specific, ASR assistance noticeably reduced the participants' delivery rate [F (1, 22) = 10.28, p = .004, η_p^2 = .319], resulted in more silent pauses [F (1, 22) = 18.63, p < .001, η_p^2 = .459], tended to increase the mean duration of silent pauses [F (1, 22) = 4.16, p = .054, η_p^2 = .159] and led to more disfluencies [F (1, 22) = 3.35, p = .081, η_p^2 = .132]. No significant interaction effect was observed for these four indicators.

6.1.4 Overall quality

No significant main effect or interaction effect was found for the participants' overall CI performance.

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

Table 3

Descriptive statistics (means with SD in brackets) for different indicators of interpreting performance

n = 23	Fidelity	Target language quality	Delivery fluency				Overall quality
			Delivery rate (wpm)	Frequency of silent pauses	Mean length of silent pauses	Disfluencies	
PAE-no ASR	6.07 (1.51)	6.64 (.98)	228.39 (55.64)	38.26 (9.61)	0.84 s (.21)	8.09 (4.03)	6.23 (1.16)
PAE-ASR	6.86 (1.26)	6.39 (.91)	201.74 (38.48)	46.91 (13.18)	1.00 s (.46)	10.78 (8.25)	6.46 (1.09)
GAE-no ASR	5.86 (.79)	6.52 (.63)	213.57 (41.17)	37.22 (10.28)	0.94 s (.40)	8.83 (6.22)	5.96 (.74)
GAE-ASR	6.90 (1.08)	6.74 (.87)	202.13 (40.01)	45.26 (11.89)	1.06 s (.47)	9.17 (5.92)	6.52 (1.23)

Note: PAE: Portuguese-accented English; GAE: General American English; ASR: automatic speech recognition; wpm: words per minute

Table 4

Results of ANOVAs for different indicators of interpreting performance

n = 23 *df* = 1, 22	Fidelity		Target language quality		Delivery Fluency								Overall quality	
					Delivery rate		Frequency of silent pauses		Mean length of silent pauses		Disfluencies			
	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2	F	η_p^2
Accent	.324	.014	.946	.041	2.134	.088	.401	.018	1.827	.077	0.288	0.013	.526	.023
ASR	7.184[*]	.246	.005	.000	10.284[**]	.319	18.630[***]	.459	4.162[△]	.159	3.346[△]	0.132	1.667	.070
Accent×ASR	.627	.028	4.676[*]	.175	2.333	.096	.029	.001	.193	.009	2.295	0.094	2.582	.105

Note: (1) ASR: automatic speech recognition; (2) [△].05 < *p* < .1; [*] *p* < .05; [**] *p* < .01; [***] *p* < .001

6.2 Cognitive load

The descriptive statistics (means and SDs) for different indicators of the participants' cognitive load are summarized in Table 5 and the results of repeated-measures ANOVAs are reported on in Table 6. No significant main effect of accent or ASR assistance, or any interaction effect between the two independent variables, was found.

Table 5

Descriptive statistics (means with SD in brackets) for different indicators of cognitive load

n = 23	Source language input phase		Target language output phase	
	Self-rated cognitive load from listening comprehension	Self-rated cognitive load from note-taking	Self-rated cognitive load from target language production	Mean *F0* from target language production
PAE-no ASR	5.67 (1.27)	5.63 (1.50)	5.88 (1.39)	188.36 (40.55)
PAE-ASR	5.62 (1.28)	5.79 (1.62)	5.83 (1.49)	185.50 (35.97)
GAE-no ASR	5.62 (.92)	5.75 (1.23)	6.13 (1.26)	188.60 (35.97)
GAE-ASR	5.83 (1.27)	6.00 (1.41)	6.04 (1.27)	186.58 (36.33)

Note: PAE: Portuguese-accented English; GAE: General American English; ASR: automatic speech recognition

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

Table 6

Results of ANOVAs for different indicators of cognitive load

n = 23 df = 1, 22	Source language input phase						Target language output phase					
	Self-rated cognitive load from listening comprehension			Self-rated cognitive load from note-taking			Self-rated cognitive Load from target language production			Mean *F0* from target language production		
	F	P	η_p^2	F	p	η_p^2	F	P	η_p^2	F	p	η_p^2
Accent	.250	.622	.011	.630	.435	.027	1.349	.257	.055	.267	.611	.012
ASR	.086	.771	.004	.648	.429	.027	.068	.796	.003	2.307	.143	.095
Accent×ASR	.476	.497	.020	.030	.864	.001	.013	.911	.001	.054	.818	.002

Note: ASR: automatic speech recognition

7. Discussion

The aim of this study was to examine the impact of ASR on the performance and cognitive load of consecutive interpreters, particularly under two conditions: with and without a challenging accent on the part of the speaker. Their CI performance was evaluated using four metrics: fidelity, TL quality, fluency, and overall quality. Fluency was specifically quantified using four sub-metrics: delivery rate, frequency and mean duration of silent pauses, and the overall occurrence of disfluencies. Cognitive load was assessed separately for the two phases of the CI process: listening and note-taking, and target-speech production.

7.1 Consecutive interpreting performance

In line with our hypothesis, ASR was found to increase fidelity during CI for both accented and non-accented speeches. This suggests that interpreters were able to produce more accurate and complete information when real-time transcription of source speeches was available, aligning with Ünlü's (2023) findings on ASR's role in improving accuracy. However, no interaction effect between ASR assistance and accent was found for fidelity, which implies that any fidelity gains afforded by ASR were relatively stable, regardless of speaker intelligibility.

The impact of ASR on TL quality for CI reveals a more complex picture. On the one hand, a significant decrease in TL quality was found when ASR was used for accented speeches, whereas, on the other, no effect was noted in the non-accented condition. This may have been due to interpreters' relying more heavily on ASR-generated transcripts when encountering unfamiliar and challenging accents, therefore adopting a hybrid approach of CI and sight translation. Some of the participants even abandoned note-taking entirely when they considered their transcription accuracy satisfactory, relying solely on the transcripts for sight translation during TL production. Such behaviour probably increased the risk of syntactic interference from the SL, as the presence of transcripts may have encouraged a more literal rendering and hindered proper deverbalization. This is particularly problematic for language pairs with significant syntactic differences, such as English and Chinese, where effective deverbalization is essential to achieving high-quality interpretation (Huang, 2013). In contrast, when interpreting the non-accented speeches, the interpreters depended less on ASR support

and more on their notes, using ASR only as a supplementary resource, therefore maintaining more consistent language quality.

Regarding fluency, ASR appeared to exert a negative influence. In particular, the use of ASR resulted in a slower delivery rate and also more frequent and longer silent pauses. A marginal increase in the overall frequency of disfluency markers was also observed with ASR use, partially corroborating Ünlü's (2003) findings which suggest that ASR increases disfluency markers. These effects may reflect divided attention during target speech production: interpreters had to switch visual focus between their notes and the on-screen transcript, or to pause note-taking momentarily to locate specific elements in the transcript. Such behaviours possibly introduced micro-hesitations and interruptions, and therefore affected the fluency metrics. Interestingly, accent did not significantly affect fluency and did not interact with ASR use. This suggests that the observed disfluency costs could be primarily attributed to the use of ASR itself rather than accent-induced listening and analysis challenges.

No statistically significant impact of ASR was observed on the overall quality of the CI output, and there was no interaction between ASR and accent on this metric. This may not come entirely as a surprise, as the improvements in fidelity brought about by ASR may have been offset by the negative impact on fluency. Previous research has shown that fluency plays a key role in the way interpreting quality is perceived, with clients often prioritizing smooth delivery (Macías, 2006). In fact, perceived fluency is often linked to users' assessments of accuracy (Rennert, 2010). For this reason, when experts subjectively assess overall quality, they may subconsciously "penalize" disfluent renditions despite a high level of information completeness and accuracy. This underscores the importance of balancing accuracy and delivery smoothness in interpreting training.

7.2 Cognitive load

Contrary to our initial hypotheses, the study found neither a significant main effect of ASR assistance on cognitive load nor any interaction effects with accent. This suggests that ASR assistance did not significantly alleviate the cognitive load of the student interpreters.

Whereas ASR assistance might offer possible benefits by supporting speech comprehension and reducing memory demands, these advantages may have been offset by the increased cognitive effort required to manage the additional input and the coordination tasks. During the listening and note-taking phase, interpreters must decide when to read the ASR text, what to note, and how to divide their visual attention between the screen (for text-reading) and paper (for note-taking) – decisions that can increase the extraneous cognitive load. Similarly, during the production phase, the interpreters often needed to refer to both their notes and the ASR transcript while producing the target speech. This dual visual–verbal processing requires constant shifts in attention which increase cognitive demands, particularly as they must also resist syntactic priming from the SL. These findings support Wickens' (2002) multiple resource theory, which posits that concurrent processing across modalities can lead to interference if attentional resources are not well coordinated. Our post-training interviews

support this interpretation, with four out of five participants expressing confusion about effectively integrating reading the ASR transcript with note-taking, and balancing reading both the ASR transcript and the self-written notes during TL production. These observations indicate a need for new instructional methods in interpreter training programmes that are specifically designed to accommodate ASR-assisted interpreting. Current training paradigms – largely developed for traditional CI – may not sufficiently resolve the unique demands posed by multimodal input. Hence, interpreting training should incorporate explicit training in ASR coordination, including selective reading strategies, eye-movement control, and note-taking techniques adapted for ASR-supported contexts.

Interestingly, no significant effect of accent was found on any of the CI performance indicators or on cognitive load. This result contrasts with much of the existing literature on the impact of accent on interpreting, which has mainly focused on SI (Albl-Mikasa, 2020; Lin et al., 2013). Several factors may explain the lack of accent effects in our study. First, although the accented speaker selected for this study was rated as challenging by experts, the speeches were nevertheless relatively understandable. Perhaps selecting speakers with even stronger accents could yield more pronounced results. But this could also reduce the accuracy of ASR-generated text, introducing confounding variables. Second, the presence of ASR could partially have mitigated accent difficulty by offering a visual backup when auditory comprehension required more effort. Third, compared to SI during which elevated time pressure exacerbates processing difficulties, CI offers more processing time, possibly enabling interpreters to become accustomed to accented input before reformulating it. It is possible that while interpreters may initially find unfamiliar accents challenging, their advanced interpreter training probably helps them to adapt more quickly. The absence of a significant effect of accent on CI performance does, however, raise questions about the extent to which accent really does affect CI performance. Further empirical research is needed to explore this matter in greater depth.

8. Conclusion

This study examined the impact of AI-powered ASR technology on CI, particularly when consecutive interpreters are faced with speeches that feature a variety of accents. A dual approach was employed, assessing both CI performance and cognitive load to provide a comprehensive picture.

Our findings show that ASR support improved interpreting fidelity but diminished fluency, which manifested as slower delivery rates and an increase in silent pauses in the output. No significant influence of ASR on TL quality was found; however, an interaction effect with accent suggests that interpreters may over-rely on ASR-generated transcripts when encountering challenging accents. ASR assistance had no significant effect on overall quality, nor was there any interaction with accent on this metric. It is possible that the improvements in fidelity offered by ASR were offset by its negative impact on fluency.

The results also reveal that ASR did not affect the cognitive load of consecutive interpreters across both phases of listening and note-taking, and production. This may be due to an

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

intricate interplay: on the one hand, ASR may alleviate cognitive pressure as interpreters take advantage of transcripts; on the other, the extra visual input may introduce additional demands on interpreters.

Whereas accent is generally rated as a challenging factor for interpreters, our findings have shown no significant impact of accent on either interpreting performance or cognitive load during CI. This calls for further in-depth empirical studies on the impact of accents in CI.

Several limitations of this study should be noted. First, only advanced interpreting students were recruited and it remains to be investigated whether these findings could be generalized to professional interpreters. Second, the present study focused on only two accent types; therefore, examining a wider range of accents with different levels of familiarity could provide a more comprehensive understanding of the impact of ASR on CI. Third, to maintain ecological validity, we relied on subjective ratings of perceived cognitive load during the listening and note-taking phase, which may not be sufficient to reveal the full picture. Nor were we able to evaluate the potential impact of the visual presentation of the ASR transcripts on interpreters' attention distribution without disrupting the natural CI process. Future research could explore innovative methodologies to deal with these challenges and possibly to provide deeper insights. Fourth, the holistic evaluations were conducted solely by expert raters, which may limit the validity of the results, since general audience perceptions might differ. Despite these limitations, our findings suggest that accessible ASR technology could enhance certain aspects of interpreting performance during CI tasks. However, its impact on cognitive load may be less beneficial than expected.

Funding information

Acknowledgement

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

References

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting*, *6*(1), 43–67. https://doi.org/10.1075/intp.6.1.05agr

Albl-Mikasa, M. (2010). Global English and English as a Lingua Franca (ELF): Implications for the interpreting profession. *Trans-kom: Zeitschrift für Translationswissenschaft und Fachkommunikation*, *3*(2), 126–148.

Albl-Mikasa, M. (2013). Express-ability in ELF communication. *Journal of English as a Lingua Franca*, *2*(1), 101–122. https://doi.org/10.1515/jelf-2013-0005

Albl-Mikasa, M. (2022). English as a lingua franca: A paradigm shift for translation and interpreting. *Slovo.ru: Baltic Accent*, *13*(1), 65–81. https://doi.org/10.5922/2225-5346-2022-1-4

Albl-Mikasa, M., Ehrensberger-Dow, M., Hunziker Heeb, A., Lehr, C., Boos, M., Kobi, M., Jäncke, L., & Elmer, S. (2020). Cognitive load in relation to non-standard language input: Insights from interpreting, translation and neuropsychology. *Translation, Cognition & Behavior*, *3*(2), 263–286. https://doi.org/10.1075/tcb.00044.alb

Albl-Mikasa, M., Fontana, G., Fuchs, L. M., Stüdeli, L. M., & Zaugg, A. (2017). Professional translations of non-native English: 'Before and after' texts from the European Parliament's Editing Unit. *The Translator*, *23*, 371–387. https://doi.org/10.1080/13556509.2017.1385940

Basel, E. (2002). *English as lingua franca: Non-native elocution in international communication. A case study of information transfer in simultaneous interpretation* [Unpublished doctoral dissertation]. University of Vienna.

Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer (version 6.3.09) [Computer software]. http://www.praat.org/

Boyer, S., Paubel, P. V., Ruiz, R., El Yagoubi, R., & Daurat, A. (2018). Human voice as a measure of mental load level. *Journal of Speech, Language, and Hearing Research*, *61*(11), 2722–2734. https://doi.org/10.1044/2018_JSLHR-S-18-0066

Chang, C., & Wu, M. M. (2014). Non-native English at international conferences: Perspectives from Chinese-English conference interpreters in Taiwan. *Interpreting*, *16*(2), 169–190. https://doi.org/10.1075/intp.16.2.02cha

Chen, S., & Kruger, J. L. (2023). The effectiveness of computer-assisted Interpreting: A preliminary study based on English–Chinese consecutive interpreting. *Translation and Interpreting Studies*, *18*(3), 399–420. https://doi.org/10.1075/tis.21036.che

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.

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. 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. https://doi.org/10.5281/zenodo.1493281

Fantinuoli, C. (2017). Speech recognition in the interpreter workstation. In J. Esteves-Ferreira, J. Macan, R. Mitkov, & O. M. Stefanov (Eds.), *Translating and the computer 39: Proceedings* (pp. 25–34). Editions Tradulex.

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

Fantinuoli, C. (2023). Towards AI-enhanced computer-assisted interpreting. In G. Corpas Pastor & B. Defrancq (Eds.), *Interpreting technologies: Current and future trends* (pp. 46–71). John Benjamins. https://doi.org/10.1075/ivitra.37.03fan

Fantinuoli, C., Marchesini, G., Landan, D., & Horak, L. (2022). KUDO interpreter assist: Automated real-time support for remote interpretation. *arXiv preprint* arXiv:2201.01800. https://doi.org/10.48 550/arXiv.2201.01800

Fantinuoli C., & Montecchio, M. (2023). Defining maximum acceptable latency of AI-enhanced CAI tools. In Ó. Ferreiro-Vázquez, A. T. V. M. Pereira, & S. L. G. Araújo (Eds.), *Technological innovation put to the service of language learning, translation and interpreting: Insights from academic and professional contexts* (pp. 213–226). Peter Lang.

Frittella, F. M., & Rodríguez, S. (2022). Putting SmarTerp to test: A tool for the challenges of remote interpreting. *INContext: Studies in Translation and Interculturalism*, *2*(2), 137–166. https://doi. org/10.54754/incontext.v2i2.21

Gaber, M., Corpas Pastor, G., & Omer, A. (2020). La tecnología habla-texto como herramienta de documentación para intérpretes: Nuevo método para compilar un corpus ad hoc y extraer terminología a partir de discursos orales en vídeo. *TRANS. Revista de Traductología*, *24*, 263–281. https://doi.org/10.24310/TRANS.2020.v0i24.7876

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(1), 65–87. https://doi.org/10.1111/j.1467-1770.1984.tb00996.x

Ghangam, S., Whitenack, D., & Nemecek, J. (2021). Dyn-ASR: Compact, multilingual speech recognition via spoken language and accent identification. *arXiv preprint* arXiv:2108.02034. https://doi.org/ 10.48550/arXiv.2108.02034

Glasser, A. (2019). Automatic speech recognition services: Deaf and hard-of-hearing usability. *Extended abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi. org/10.1145/3290607.3308461

Gile, D. (1995). Fidelity assessment in consecutive interpretation: An experiment. *Target*, 7(1), 151–164. https://doi.org/10.1075/target.7.1.12gil

Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (rev. ed.). John Benjamins. https://doi.org/10.1075/btl.8

Goldsmith, J., & Blouin, L. P. (2021). *Use automatic speech recognition to improve consecutive interpreting: Sight consec* [Video]. Youtube. https://youtu.be/f1X0i5S0cuE

Gutz, S. E., Stipancic, K. L., Yunusova, Y., Berry, J. D., & Green, J. R. (2022). Validity of off-the-shelf automatic speech recognition for assessing speech intelligibility and speech severity in speakers with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, *65*(6), 2128–2143. https://doi.org/10.1044/2022_JSLHR-21-00589

Hacking, C., Verbeek, H., Hamers, J. P. H., & Aarts, S. (2023). The development of an automatic speech recognition model using interview data from long-term care for older adults. *Journal of the American Medical Informatics Association*, *30*(3), 411–417. https://doi.org/10.1093/jamia/oc ac241

Hamidi, M., & Pöchhacker, F. (2007). Simultaneous consecutive interpreting: A new technique put to the test. *Meta*, *52*(2), 276–289. https://doi.org/10.7202/016070ar

Han, C., Chen, S., Fu, R., & Fan, Q. (2020). Modelling the relationship between utterance fluency and raters' perceived fluency of consecutive interpreting. *Interpreting*, *22*(2), 211–237. https://doi.

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

org/10.1075/intp.00040.han

Han, C., Xiao, R., & Su, W. (2021). Assessing the fidelity of consecutive interpreting. *Interpreting*, *23*(2), 245–268. https://doi.org/10.1075/intp.00058.han

Huang, Y. (2013). *Sentence memory and deverbalization* [Doctoral dissertation, Shanghai International Studies University]. China National Knowledge Infrastructure.

Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., & Leino, T. (2011). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Applied Ergonomics*, *42*(2), 348–357. https://doi.org/10.1016/j.apergo.2010.08.005

Jensen, C., & Thøgersen, J. (2017). Foreign accent, cognitive load and intelligibility of EMI lectures. *Nordic Journal of English Studies*, *16*(3), 107–137. https://doi.org/10.35360/njes.414

Johannes, B., Wittels, P., Enne, R., Eisinger, R., Castro, C. A., Thomas, J. L., Adler, A. B., & Gerzer, R. (2007). Non-linear function model of voice pitch dependency on physical and mental load. *European Journal of Applied Physiology*, *101*(3), 267–276. https://doi.org/10.1007/s00421-007-0496-6

Katikos, R. (2015). *Simultaneous interpreting and English as a lingua franca in view of the interlanguage speech intelligibility benefit* [Unpublished master's thesis]. University of Vienna.

Kurz, I. (2008). The impact of non-native English on students' interpreting performance. In G. Hansen, A. Chesterman, & H. Gerzymisch-Arbogast (Eds.), *Efforts and models in interpreting and translation research* (pp. 179–192). John Benjamins. https://doi.org/10.1075/btl.80.15kur

Kurz, I., & Basel, E. (2009). The impact of non-native English on information transfer in simultaneous interpretation. *Forum*, *7*(2), 187–212. https://doi.org/10.1075/forum.7.2.08kur

Levis, J. M., & Zhou, Z. (2018). Accent. In J. I. Liontas, & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching* (pp. 1–5). John Wiley & Sons. https://doi.org/10.1002/9781118 784235.eelt0002

Li, T., & Chmiel, A. (2024). Automatic subtitles increase accuracy and decrease cognitive load in simultaneous interpreting. *Interpreting*, *26*(2), 253–281. https://doi.org/10.1075/intp.00111.li

Li, Z., Miao, H., Deng, K., Cheng, G., Tian, S., Li, T., & Yan, Y. (2021). Improving streaming end-to-end ASR on transformer-based causal models with encoder states revision strategies. *arXiv preprint* arXiv:2207.02495. https://doi.org/10.48550/arXiv.2207.02495

Lin, I. I., Chang, F. A., & Kuo, F. (2013). The impact of non-native accented English on rendition accuracy in simultaneous interpreting. *The International Journal of Translation and Interpreting Research*, *5*(2), 30–44. https://doi.org/10.12807/ti.105202.2013.a03

Lippi-Green, R. (2011). *English with an accent: Language, ideology, and discrimination in the United States*. Routledge. https://doi.org/10.4324/9780203348802

Mackintosh, J. (2003). The AIIC workload study. *Forum*, *1*(2), 189–214. https://doi.org/10.1075/forum.1.2.09mac

Macías, M. P. (2006). Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency. *Interpreting*, *8*(1), 25–42. https://doi.org/10.1075/intp.8.1.03pra

McAllister, R. (2000). Perceptual foreign accent and its relevance for simultaneous interpreting. In B. E. Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting: Interdisciplinary perspectives* (pp. 45–64). John Benjamins. https://doi.org/10.1075/btl.40.05mca

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign accented speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

Mustaquim, M. M. (2013). Automatic speech recognition: An approach for designing inclusive games. *Multimed Tools and Applications*, 66(1), 131–146. https://doi.org/10.1007/s11042-011-0918-7

Orlando, M. (2015). Digital pen technology and interpreter training, practice, and research: Status and trends. In S. Ehrlich & J. Napier (Eds.), *Interpreter education in the digital age: Innovation, access, and change* (pp. 125–152). Gallaudet University Press. https://doi.org/10.2307/j.ctv2rcnmhs.11

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Park, J. S., & Na, H. J. (2020). Front-end of vehicle-embedded speech recognition for voice-driven multi-UAVs control. *Applied Sciences*, 10(19), Article 6876. https://doi.org/10.3390/app10196876

Pérez-Ramón, R., Lecumberri, M., & Cooke, M. (2022). Foreign accent strength and intelligibility at the segmental level. *Speech Communication*, *137*, 70–76. https://doi.org/10.1016/j.specom.2022.01.005

Pisani, E., & Fantinuoli, C. (2021). Measuring the impact of automatic speech recognition on number rendition in simultaneous interpreting. In B. Zheng & C. Wang (Eds.), *Empirical studies of translation and interpreting: The post-structuralist approach* (pp. 181–197). Routledge. https://doi.org/10.4324/9781003017400-14

Pöchhacker, F. (2016). *Introducing interpreting studies* (2nd ed.). Routledge. https://doi.org/10.4324/9781315649573

Prandi, B. (2018). An exploratory study on CAI tools in simultaneous interpreting: Theoretical framework and stimulus validation. In C. Fantinuoli (Ed.), *Interpreting and technology* (pp. 29–59). Language Science Press. https://doi.org/10.5281/zenodo.1493281

Prandi, B. (2023). *Computer-assisted simultaneous interpreting: A cognitive-experimental study on terminology*. Language Science Press. https://doi.org/10.5281/zenodo.7143055

Rennert, S. (2010). The impact of fluency on the subjective assessment of interpreting quality. *The Interpreters' Newsletter*, *15*, 101–115.

Romero, M., Gómez-Canaval, S., & Torre, I. G. (2024). Automatic speech recognition advancements for indigenous languages of the Americas. *Applied Sciences*, *14*(15), Article 6497. https://doi.org/10.3390/app14156497

Sabatini, E. (2000). Listening comprehension, shadowing and simultaneous interpretation of two "non-standard" English speeches. *Interpreting*, *5*(1), 25–48. https://doi.org/10.1075/intp.5.1.03sab

Seresi, M., & Láncos, P. L. (2022). Teamwork in the virtual booth: Conference interpreters' experiences with RSI platforms. In K. Liu & A. K. F. Cheung (Eds.) *Translation and interpreting in the age of COVID-19* (pp. 181–196). Springer. https://doi.org/10.1007/978-981-19-6680-4_10

Su, W., & Li, D. (2024). Cognitive load and interpretation quality of technology-assisted simultaneous interpreting. *Foreign Language Teaching and Research*, *56*(1), 125–135. https://doi.org/10.19923/j.cnki.fltr.2024.01.012

Sun, H., Li, K., & Lu, J. (2021). AI-assisted simultaneous interpreting: An experiment and its implications. *Technology Enhanced Foreign Language Education*, *6*, 75–80.

Sung, C. (2016). Does accent matter?: Investigating the relationship between accent and identity in English as a lingua franca communication. *System*, *60*, 55–65. https://doi.org/10.1016/j.system.2016.06.002

Tammasrisawat, P., & Rangponsumrit, N. (2023). The use of ASR-CAI tools and their impact on

Yu, Z. & Shao, Z. (2025). Does automatic speech recognition help in consecutive interpreting that involves varying accents? Evidence from interpreting performance and cognitive load. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 137–159.

interpreters' performance during simultaneous interpretation. *New Voices in Translation Studies, 28*(2), 25–51. https://doi.org/10.14456/nvts.2023.27

Tissi, B. (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis. *The Interpreter's Newsletter*, *10*, 103–126.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84–119. https://doi.org/10.1093/applin/17.1.84

Ünlü, C. (2023). *Automatic speech recognition in consecutive interpreter workstation: Computer-aided interpreting tool "Sight-Terp"* [Unpublished master's thesis]. Hacettepe University.

Wang, C., Wu, Y., Liu, S., Li, J., Lu, L., Ye, G., & Zhou, M. (2020). Low latency end-to-end streaming speech recognition with a scout network. *arXiv preprint* arXiv:2003.10369. https://doi.org/10.48550/arXiv.2003.10369

Wang, X., & Wang, C. (2019). Can computer-assisted interpreting tools assist interpreting? *Transletters: International Journal of Translation and Interpreting*, *3*, 109–139.

Will, M. (2020). Computer-aided interpreting (CAI) for conference interpreters: Concepts, content and prospects. *Journal for Communication Studies*, *13(*1[25]), 37–71.

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*, Article 100131. https://doi.org/10.1016/j.amper.2023.100131

Zhu, J., Wang, D., & Zhao, Y. (2024). Design of smart home environment based on wireless sensor system and artificial speech recognition. *Measurement: Sensors*, *33*, 101090. https://doi.org/10.1016/j.measen.2024.101090