

RESEARCH ARTICLE

# Simultaneous interpreting with auto-subtitling: Investigating viewer cognitive effort, stress, and comprehension

Yanlin Li<sup>1</sup>, Jiawen Diao<sup>2</sup>, Andrew K. F. Cheung<sup>2\*</sup>

**1** Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong, **2** Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Kowloon, Hong Kong

\* [andrew.cheung@polyu.edu.hk](mailto:andrew.cheung@polyu.edu.hk)



**OPEN ACCESS**

**Citation:** Li Y, Diao J, Cheung AKF (2025) Simultaneous interpreting with auto-subtitling: Investigating viewer cognitive effort, stress, and comprehension. PLoS One 20(8): e0330692. <https://doi.org/10.1371/journal.pone.0330692>

**Editor:** Robyn Berghoff, Stellenbosch University, SOUTH AFRICA

**Received:** February 2, 2025

**Accepted:** August 4, 2025

**Published:** August 22, 2025

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0330692>

**Copyright:** © 2025 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Abstract

Simultaneous interpreting (SI) enables real-time cross-language communication without significant delays and is vital for fast-paced environments such as multilingual conferences. Automatic subtitles, powered by artificial intelligence (AI), is an important mode of audiovisual translation and has been widely deployed by virtual conferencing platforms to help users overcome language barriers. While the cognitive and emotional impacts of SI have been explored in prior studies, research directly comparing SI, auto-subtitling, and their combined use remains limited. This study investigates and compares the effectiveness of three interlingual translation modes, auto-subtitling, SI, and a combined dual-modality approach, on comprehension, cognitive load, and stress levels. Mandarin Chinese-speaking participants viewed a video presentation delivered in Arabic, a language they did not understand. Participants were divided into three groups: Group A relied on automatic subtitles in Simplified Chinese characters, Group B relied on SI in Mandarin, and Group C used a combination of both methods. Analysis of electroencephalographic data and comprehension test results revealed no statistically significant differences in content comprehension across the groups. However, Group A reported the poorest viewing experience, with the highest stress levels, while Group B expended the greatest cognitive effort. Group C exhibited the lowest levels of cognitive effort and stress, underscoring the advantages of dual-modality systems. These findings suggest that combining accurate automatic subtitles with professional interpreting may enhance accessibility, reduce cognitive demands, and improve the viewing experience, offering valuable insights into the integration of AI-driven technologies in SI.

## Introduction

The purpose of this research is to experimentally examine the effects of three interlingual modes of translation, i.e., auto-subtitling, simultaneous interpreting (SI), and

**Data availability statement:** All relevant data are within the manuscript and its [Supporting information](#) files.

**Funding:** We would like to thank the Research and Innovation Office of the Hong Kong Polytechnic University for supporting the project (Project Code: RKQY). This manuscript was also partly supported by funding from Project number P0046386 of the Department of Chinese and Bilingual Studies of the Hong Kong Polytechnic University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

a combination of both (dual-modality), on viewer comprehension, cognitive load, and viewing experience during multilingual presentations. Specifically, the study evaluates whether automatic subtitles, generated by artificial intelligence (AI), can facilitate content comprehension comparably to SI, analyzes the cognitive demands experienced by viewers across these three conditions, and identifies the most effective mode for cross-linguistic communication. By systematically addressing these aspects, this research seeks to enhance audience understanding and engagement in multilingual settings.

The rise of multimedia communication has driven the development of audiovisual translation (AVT), encompassing subtitling, dubbing, and voiceover. Among these, subtitling is one of the most widely adopted due to its cost-effectiveness and efficiency [1,2]. Originally developed to facilitate access for deaf/hard-of-hearing audiences and those with limited proficiency in the source language, subtitling now serves critical functions in both educational and entertainment contexts [3–5]. In language education, subtitles have been shown to support vocabulary acquisition, listening comprehension, and learner engagement by providing concurrent visual and auditory input [6]. Research further highlights their role in long-term language development, particularly through autonomous exposure to captioned media (e.g., films, TV programmes) in target languages [7]. In the entertainment industry, subtitling enables cross-cultural access to audiovisual narratives, allowing global audiences to engage with original-language performances, a function amplified by streaming platforms like Netflix, YouTube, and TikTok [8].

Traditionally, real-time subtitling relied on stenotyping or skilled subtitlers using dual keyboards [1]. These labor-intensive methods limited scalability until the advent of AI-driven solutions. Automatic speech recognition (ASR) has transformed subtitling practices. ASR, which converts spoken language into written text using algorithms, has enabled the adoption of real-time automatic subtitles, also known as live captioning [9]. These systems are embedded in platforms like Zoom and Microsoft Teams, providing near-instantaneous subtitles to enhance accessibility for global audiences [8]. While ASR technologies now enable real-time automatic subtitles in general contexts, challenges persist in specialized, high-stakes settings, such as multilingual conferences, where SI is traditionally employed.

SI is a cornerstone of multilingual communication, enabling real-time exchange across linguistic boundaries in international and institutional contexts [1,2]. This mode of interpretation relies on highly skilled human interpreters who deliver accurate and immediate renditions of the source language into the target language [9]. While effective, professional SI services are resource-intensive, requiring significant human and financial investment [10]. The pandemic-driven shift to virtual and hybrid conferencing has amplified these challenges, exposing the limitations of traditional SI services in terms of scalability and affordability [11,12]. Although remote SI has emerged as a solution, it remains constrained by interpreter availability and cost. Recent advancements in auto-subtitling offer a complementary alternative. These systems provide real-time textual support to enhance accessibility in formal, multilingual conference settings, such as academic symposia and international policy forums

[10]. AI-powered tools like Otter.ai and Google Meet's live captioning are increasingly utilized to support multilingual communication in these contexts.

Existing research has extensively examined the advantages and limitations of subtitles across various contexts. Automatic subtitles, created using ASR technology, have been shown to enhance comprehension by bridging auditory and textual input, improving immersion, and maintaining viewer focus [13–17]. However, such advantages are not universal. When subtitle speed is high, content is dense, or synchronization is poor, the sequential processing of visual (textual) and auditory information may actually increase cognitive effort, potentially undermining comprehension [18,19].

This study focuses specifically on open (burned-in) interlingual automatic subtitles, machine-generated translations embedded directly into video, used in real-time or pre-recorded multilingual communication. A proposed enhancement in such settings is the dual-modality approach, which combines SI with auto-subtitling to distribute information across both auditory and visual channels [16]. While this multimodal design holds promise for supporting diverse processing preferences and compensating for gaps in either channel, it may also impose greater cognitive demands, as viewers are required to shift attention between listening and reading, particularly in high-stakes or fast-paced scenarios [11,20]. While auto-subtitling systems have improved, errors in transcription or translation persist, particularly in noisy audio or linguistically complex content [20].

While studies have explored the integration of ASR tools in interpreting workflows, they primarily focus on the impact on interpreters' performance rather than the viewers' experience with dual-modality systems [9]. For instance, Defrancq and Fantinuoli [21] highlight the psychological benefits of ASR availability for interpreters, while Tammasrisawat and Rangponsumrit [22] demonstrate that ASR-computer-assisted interpreting (CAI) tools can reduce error rates and omissions, improving interpreters' terminology rendition quality. However, little is known about how the combination of automatic subtitles and SI affects viewers' comprehension, cognitive load, and overall experience. It remains unclear whether this approach enhances understanding or introduces new challenges due to the cognitive demands of processing dual modalities.

This gap in the literature underscores the need for a rigorous investigation into the combined effect of auto-subtitling and SI on viewers. Such research can clarify whether the dual-modality approach strikes the right balance between accessibility and cognitive demands, offering practical insights for improving multilingual communication. To address this gap, we propose the following research questions (RQs):

**RQ1:** How do automatic subtitles compare to SI in facilitating content comprehension for viewers?

**RQ2:** What differences in cognitive demands do automatic subtitles and SI impose on viewers?

**RQ3:** How do automatic subtitles and SI influence viewers' overall viewing experience?

The contributions of this study are twofold. First, it examines how audiences cognitively respond to dual modality input, which is the combination of SI and automatic subtitles, by evaluating comprehension, cognitive effort, and viewing experience in multilingual conference settings where real-time processing is essential. While prior research has focused on interpreters' performance or single-modality delivery, this study addresses a notable gap by providing empirical analysis of audience experience in dual-modality environments, offering insights in AVT. Second, the study enhances methodological rigor by employing electroencephalographic (EEG) techniques to objectively measure viewers' cognitive load. This approach moves beyond self-reported data to reveal how cognitive resources are allocated during dual-modality processing.

## Literature review

### Automatic subtitles and the viewing experience

Traditional subtitles serve as textual representations of spoken information in videos, fulfilling a variety of essential functions. Initially created to provide translations by human translators for movies and television audiences facing language barriers, subtitles have evolved into indispensable tools for language learning, cognitive load management, and improving viewer comprehension [18,23–25].

Automatic subtitles are generated by ASR technology that enables the transcription of spoken language into text by analyzing the shape of speech waves [26]. Unlike traditional subtitles, automatic subtitles are generated in real time, adapting to the speaker's pace and often employing innovative formats such as multi-line stacking to improve readability [27]. Auto-subtitling has gained traction for its efficiency and scalability, particularly in environments requiring immediate content delivery, such as live broadcasts, online learning platforms, and virtual meetings [21].

ASR has undergone significant advancements since its inception. In 1952, Davis et al. [28] pioneered early speech recognition systems capable of recognizing digits, achieving notable accuracy despite the technological limitations of the time. Modern ASR systems, powered by AI and neural network architectures such as the Transformer model, have since revolutionized speech-to-text technology, enabling real-time transcription of large-scale audio data with greater accuracy and reliability [29]. However, accuracy in ASR remains a multifaceted challenge, particularly in real-world environments. This includes challenges in recognizing speech accurately in the same language (monolingual transcription) and auto-translating speech into another language (interlingual transcription) [9,16,17,30]. Factors such as background noise, accents, dialects, and linguistic variations further complicate recognition tasks [29,30]. These limitations are particularly critical in high-stakes settings such as live conferences, where errors in recognition or translation could lead to miscommunication or misunderstandings.

Despite these challenges, automatic subtitles have become an essential tool for facilitating accessibility. They are especially valuable for ensuring accessibility for deaf or hard of hearing individuals, as well as for native [31] and non-native speakers [32], including those in noisy or quiet environments, multitaskers, and others with specific needs or preferences. Platforms such as Zoom and Microsoft Teams have integrated auto-subtitling features as part of their services to address accessibility needs. Although these features are typically included in paid licenses, they are perceived as lower-cost alternatives to human-provided services, such as SI or professional subtitling, due to the scalability and automation of ASR technology [20]. Auto-subtitling might not be as effective as having SI because it lacks the nuances of spoken language that help keep people interested and oriented [33]. Therefore, some may prefer overcoming language barriers by listening to SI because the need to expend excessive cognitive effort to follow visually a narrative may result in a negative viewing experience [34].

Research to date has focused on the impact of automatic subtitles on interpreters themselves [9,21,22]. These discussions have largely emphasized how auto-subtitling can support interpreters by improving their accuracy and reducing cognitive load [35]. However, listeners can also benefit from the integration of automatic subtitles. In virtual conferences, where there is a need to accommodate diverse linguistic requirements and ensure accessibility for a broader audience, automatic subtitles can play a crucial role in enhancing the delivery of multilingual content. This is particularly important in settings with participants from varied linguistic and cognitive backgrounds, where balancing visual and auditory inputs is essential for optimizing information accessibility [10].

### Measuring cognitive effort and viewing experience using EEG data

As auto-subtitling systems become increasingly prevalent across various contexts [3,21,22,36], understanding their impact on viewers' cognitive performance and comprehension is crucial for optimizing communication and reducing language barriers [19]. The concepts of cognitive effort and stress may appear similar but are in fact distinct. Cognitive effort is typically defined as the mental resources allocated for processing a task [36], whereas stress refers to a particular relationship between an individual and their environment in which the demands of the situation are perceived as taxing or exceeding the individual's resources, potentially affecting their focus and functional performance [37]. In the context of viewers' audiovisual content comprehension, cognitive effort relates to the mental resources required to process information when viewing automatic subtitles and simultaneously listening to spoken content [19]. Stress, on the other hand, arises when the cognitive demands of processing audiovisual information are perceived as overwhelming or exceeding an individual's comfort level [37]. This can affect viewers' ability to maintain focus and process content effectively, especially

in high-pressure environments such as live conferences. Research into viewers' cognitive performance in audiovisual content comprehension, particularly in the context of automatic subtitles, represents a vital area of study. It investigates the relationship between multimodal content processing and cognitive functions, providing valuable insights into how technologies like auto-subtitling can either support or obstruct comprehension, ultimately shaping the effectiveness of information delivery.

To effectively measure cognitive effort and stress, researchers commonly utilize both subjective and objective assessment methods. Subjective measures, such as self-report questionnaires, enable participants to describe their perceived levels of cognitive effort and stress during specific tasks [38,39]. For instance, self-reported data in [40] suggested that computer-assisted consecutive interpreting resulted in fewer pauses, reduced cognitive load, and improved overall interpreting quality, particularly in terms of accuracy. However, conflicting findings have emerged. The conflict matrix, which lacks "reliable empirical data to substantiate" its claims [41], suggests that while additional information might logically enhance interpretation accuracy, comparisons between cognitive resource allocation and the conflict matrix for SI with and without text predict a discrete increase in task interference, and consequently, cognitive load. These conflicting findings may, in part, be attributed to the limited accuracy and reliability of subjective measures [41].

Objective measures offer a detailed understanding of cognitive effort and stress by tracking real-time physiological responses linked to cognitive processes. These responses usually involve the central and peripheral nervous systems [42]. The most common way to accurately decode the rhythm of the nervous system is to use EEG data [43], which enables researchers to correlate specific brain responses with performance metrics. In recent years, advancements in human-computer interaction and the increased accessibility of low-cost consumer-grade EEG devices have led to widespread recognition within research communities that objective EEG data provide a robust and reliable means of measuring participants' cognitive performance and the user experience [9,44–47].

Analyzing brain wave patterns is essential for understanding cognitive effort and stress levels [45]. While the relationship between alpha wave fluctuations and cognitive workload remains somewhat unclear [48], neural oscillations associated with theta and beta waves are recognized as reliable markers of cognitive workload [49]. According to previous studies [42,50], theta waves (4–8 Hz) are associated with memory processing and cognitive load; increased theta activity may indicate significant cognitive effort and greater strain on working memory [50] and increased mental fatigue [51–53]. In such a setting, theta power can indicate mental resource allocation during multimodal (audiovisual) content processing. Higher theta activity in the frontal region corresponds to executive functions, including decision-making, working memory, and semantic processing, while occipital region activity reflects visual processing and reading comprehension [54]. Regarding beta waves, Lazarus and Folkman [55] identified elevated beta wave power as an indicator of increased stress levels, as beta waves are associated with heightened cognitive activity and emotional strain during task performance. In particular, high beta wave activity often reflects mental states characterized by tension, anxiety, or demands associated with sustaining attention under pressure. These findings are further supported by research in interpreting and language translation contexts, which has used beta wave activity to measure cognitive and emotional strain [51,52].

The adoption of EEG technology has been instrumental in observing the cognitive mechanisms of simultaneous interpreters. Early studies by Kurz [56,57] explored EEG changes during SI, providing valuable insights into the mental processes underlying complex verbal tasks. These studies also confirmed the utility of computer-assisted neurophysiological measures for investigating cortical processes during SI. Building on this foundation, Koshkin et al. [58] employed EEG techniques to examine interpreters' neural activity under varying levels of working memory load during the SI of continuous prose. Similarly, Moser-Mercer [59] underscored the importance of studying cognitive functioning in simultaneous interpreters, highlighting the "plasticity" of the interpreter's brain and how experienced professionals can overcome common cognitive constraints through expertise. Yagura et al. [60] further advanced this area by quantifying differences in brain activity patterns between experienced and novice interpreters using EEGs in realistic SI environments. Their findings revealed that years of SI experience significantly influence selective attention during interpretation.

While existing studies primarily focus on the use of EEG techniques to investigate simultaneous interpreters' cognitive processes, research on viewers' cognitive performance under different interpreting modes remains sparse. Specifically, the combined effects of auto-subtitling systems and SI on viewers' cognitive performance have yet to be thoroughly explored. Given the growing emphasis on leveraging advanced technologies to enhance information accessibility, it is crucial to examine how these modes interact. This knowledge gap is particularly significant in light of the increasing integration of technology into interpreting contexts. Addressing this gap is essential not only for advancing theoretical understanding of interpreting practices but also for developing evidence-based approaches to improve interpreting performance and enhance user satisfaction in multilingual communication settings.

## Method

This study adopted a mixed-methods approach to examine participants' cognitive performance and comprehension. Participants' comprehension was assessed using an online questionnaire, while an EEG device measured their cognitive performance.

Ethics approval was obtained from The Hong Kong Polytechnic University Institutional Review Board (PolyU IRB, reference number HSEARS20241022005, 28 October 2024). Participants, recruited via offline classes, provided oral consent after reviewing a clear, non-technical information sheet outlining the study's aims and procedures. They were assured of their right to withdraw at any time without penalty.

## Participants

A total of 45 participants (28 female, 17 male) took part in the study, with 15 individuals in each of the three experimental groups (see [Table 1](#) for group-specific demographics). All participants were native speakers of Mandarin Chinese and reported using Mandarin as their primary language of communication. Self-assessed proficiency aligned with the C2 level of the Common European Framework of Reference for Languages (CEFR). Screening questions confirmed no history of visual, auditory, or language-related impairments.

Participants were born and raised in mainland China, with representation from multiple provinces and municipalities, including Beijing, Shanghai, Guangdong, and Sichuan. They were recruited from a range of academic disciplines to ensure variation in educational background and to minimize domain-specific bias. None of the participants reported any academic or professional background in fields such as international relations, diplomacy, or United Nations (UN) affairs, based on the information gathered through background questionnaires. Although participants were not directly asked about their prior familiarity with the topic of the stimulus video, their disciplinary profiles and lack of relevant training suggest limited exposure to the domain. This helps to mitigate the influence of domain expertise on task performance and ensures that the study primarily captures general audience comprehension.

To examine the effects of different content delivery methods on cognitive effort, viewing experience, and comprehension, the 45 participants were randomly assigned to three experimental groups of equal size ( $n = 15$ ): Group A (automatic subtitles), Group B (SI), and Group C (a combination of automatic subtitles and SI).

**Table 1. Demographic characteristics of participants (N = 45).**

Demographic Characteristics	Total (N=45)	Group A	Group B	Group C
Female	28	10	9	9
Male	17	5	6	6
Average Age	23.7	23.5	23.6	23.9
Age Range	21-26	21-25	21-26	21-26

<https://doi.org/10.1371/journal.pone.0330692.t001>

While stratified randomization was not employed at the assignment stage, this decision was informed by the modest sample size and logistical considerations. Given that demographic variables were not expected to exert strong effects on the dependent measures, simple randomization was deemed sufficient. Post-allocation checks confirmed satisfactory balance across key participant characteristics. Gender distribution was comparable across groups (Group A: 10F/5M, Group B: 9F/6M, Group C: 9F/6M), and participant ages ranged from 21 to 26 years, with group averages between 23.5 and 23.9 (see [Table 1](#)). Participants were also diverse in their academic backgrounds, with no concentration of disciplinary expertise in any one group. These checks support the internal validity of the group comparisons and help rule out confounding effects from participant-level variables.

The formal experiment commenced on November 1 and concluded on November 18, 2024. All participants wore EEG devices while watching the conference video in individual rooms without any intervention. We monitored their EEG signals from a separate independent room. To control for potential stress-inducing factors, the experimental environment was designed to be quiet and low-pressure. Tasks were untimed, and no performance-based feedback was provided. Participants were encouraged to complete the questionnaire at their own pace. Baseline measures of cognitive state were also collected to account for individual differences in initial stress levels. Informal debriefings indicated that participants generally did not perceive the task as stressful. These measures help ensure that any observed effects are attributable to the experimental manipulation.

### EEG headset and key brain waves

The Emotiv EPOC X 14-channel wireless EEG headset was used to record brain activity, with raw EEG data accessed through the Emotiv Pro Suite installed on an HP Pavilion Plus laptop. The headset was paired with Emotiv Pro via Bluetooth, enabling effective real-time monitoring of EEG data. The Emotiv headset provides the power of five frequency bands, namely theta (4–8 Hz), alpha (8–12 Hz), low beta L (12–16 Hz), high beta H (16–25 Hz), and gamma (25–45 Hz), for each sensor, as shown in [Fig 1](#).

### Materials

A single speech in Arabic delivered at a UN conference was selected as the experimental material. This speech was chosen because it reflects the UN's standard practice of providing SI services into its six official languages, aligning with our focus on evaluating comprehension of interpreted content in professional settings. The original speech was interpreted into Chinese by professional UN interpreters, and the official video and audio recordings were publicly available on the UN website. To ensure experimental control, all participants across the three groups were exposed to identical speech content, which was then processed into three distinct delivery modes:

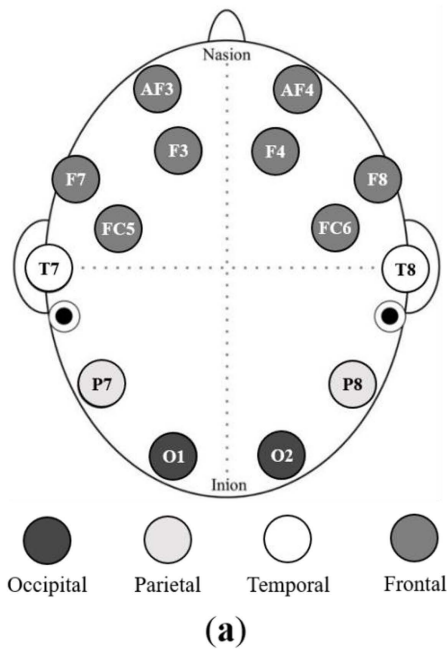
Version A: Chinese auto-subtitling generated by an ASR system based on the Chinese SI audio.

Version B: Chinese SI audio only, without subtitles.

Version C: A combination of Chinese SI audio and Chinese automatic subtitles.

The speech content, pacing, and audiovisual quality were held constant across all three versions. Each version lasted exactly 3 minutes and 40 seconds. The use of a single speech across all experimental conditions was an intentional design choice aimed at maximizing internal validity. This design ensured that the only manipulated variable was the mode of delivery, while the linguistic content, discourse topic, and communicative context remained constant. Using a single speech allowed us to isolate the effects of delivery mode on cognitive effort and comprehension outcomes without introducing variability from message content or situational features.

To assess the participants' comprehension of the speech in the target language, a multiple-choice test was developed in Simplified Chinese characters. The test consisted of 10 single-choice questions focused on factual recall, such as the



Brain Wave	Frequency	Location
Delta	0.5-3 Hz	Frontal region
Theta	4-7 Hz	Parietal and temporal regions
Alpha	8-13 Hz	Occipital region
Beta	14-30 Hz	Parietal and frontal regions

**Fig 1. Electrode placement of the Emotiv Epoc X EEG headset.** AF: anterior–frontal, F: frontal, FC: fronto-central, T: temporal, P: parietal, O: occipital. The odd numbers denote the left hemisphere and even numbers represent the right hemisphere. Black dots indicate locations of reference electrodes.

<https://doi.org/10.1371/journal.pone.0330692.g001>

speech’s main topics, numerical data, and named entities, to evaluate the participants’ immediate understanding of the content. An example of a question translated in English is as follows:

“What is the purpose of holding this summit?”

- A) To promote global economic recovery
- B) To emphasize national sovereignty and integrity
- C) To enhance economic cooperation among countries
- D) To reshape the multilateral system”

Correct answer:

D) This option directly reflects the central goal as stated in the speech: “reshape the multilateral system”.

Distractor rationale:

- A) Reflects general UN themes related to development and recovery but does not align with the summit’s explicitly stated focus.
- B) Draws on the principle of respecting sovereignty mentioned in the speech but misrepresents it as the main purpose.
- C) Suggests international cooperation, a peripheral theme, but inaccurately elevates it to the summit’s core objective.

Each question was worth 10 points, with no partial credit awarded for incorrect or incomplete answers. As a result, the total comprehension score ranged from 0 to 100, with higher scores indicating better comprehension. To ensure the test’s

validity and reliability, all questions and answer options were reviewed by a linguistics expert who specialized in translation and interpreting. This review ensured that the questions aligned with the speech content and that the wording of the answer options was clear, precise, and appropriate for assessing comprehension of the target language.

## Experimental procedure

A pilot test was conducted prior to the main study to ensure the feasibility of the experimental design and procedures. This included evaluating the functionality of video playback, the clarity of the instructions, and the timing of the comprehension test. Feedback from the pilot phase informed minor adjustments, including the rewording of comprehension questions to improve clarity and technical refinements to ensure the smooth functioning of all experimental components. One such revision involved replacing the original item “*Why was the summit important?*”, which invited subjective interpretation, with “*What is the purpose of holding this summit?*”, a fact-based question more directly anchored in the explicit linguistic content of the source text. The pilot confirmed the feasibility of the materials and procedures for the main experiment. Accordingly, the formal study adopted the validated setup—including standardized video playback procedures—and was conducted in a controlled lab environment to ensure consistency and data reliability.

Participants were randomly assigned to one of three experimental conditions—SI, automatic subtitles, or dual-modality (SI + automatic subtitles)—using a computer-generated randomization sequence. Each participant experienced only one modality condition and viewed the same stimulus video in a fixed order, followed by comprehension questions. Random allocation was employed to control for potential between-group variation unrelated to the experimental treatment.

Each participant selected a session time based on their availability and was informed in advance that they would receive a HK\$50 gift card as compensation. Based on each participant’s selected schedule, the research team reserved the lab and ensured the appropriate condition materials were prepared. To minimize bias, participants were explicitly instructed not to discuss their experiences or the test content with one another until all sessions were completed.

Upon arrival, the participant was provided with a consent form to read and sign, ensuring ethical compliance with the study protocol. Following this, the participant was briefed on the experimental procedure and introduced to the wireless EEG device, which was used to monitor their cognitive activity. Each experimental session was conducted individually. A trained research assistant assisted the participant in correctly positioning the wireless EEG headset and ensured that 100% contact quality was achieved before proceeding with the task.

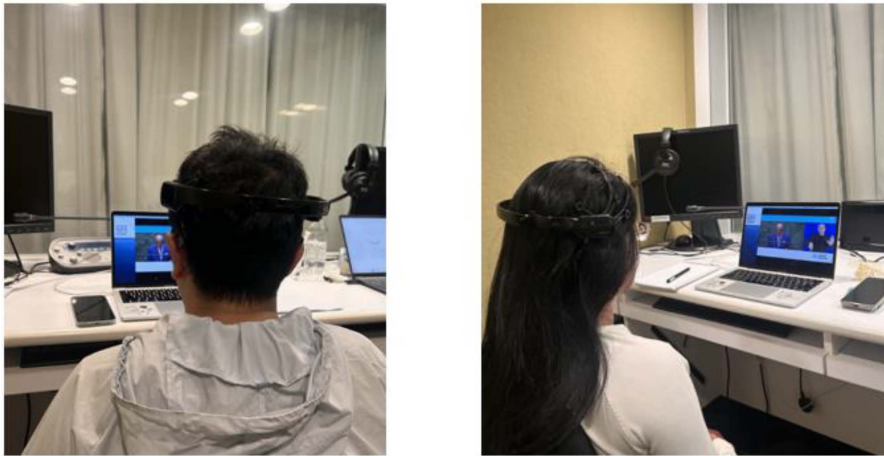
Once the EEG setup and baseline signal confirmation were complete, the participant was seated at an individual station and the video presentation commenced. During the video viewing, researchers monitored the participant from a separate room to ensure uninterrupted EEG data collection and to quickly address any technical issues. The environment was carefully controlled to eliminate distractions, such as noise or external interference, ensuring consistent testing conditions for each participant.

To maintain consistency, the video could not be paused or replayed during the experiment. The comprehension questions were only accessible after the video had ended, to ensure that the participant relied solely on the presented material. [Fig 2](#) presents the participant wearing the EEG device while watching the video material.

## Results

### Content comprehension

After all participants completed the experiments, their comprehension scores were collected and processed for statistical analysis. To compare comprehension scores across the three groups, the Kruskal-Wallis test was employed. The analysis yielded an  $H$  statistic of 0.407 ( $df=2$ ,  $p=0.816$ ). As  $p > 0.05$ , the null hypothesis of equal distributions across groups was not rejected. This result indicates that there were no statistically significant differences in comprehension scores among the three groups.



**Fig 2. Participant wearing the EEG device while watching the video.**

<https://doi.org/10.1371/journal.pone.0330692.g002>

Descriptive statistics (Table 2) show a gradual increase in mean scores: Group A (automatic subtitles) scored 70.00, Group B (SI) scored 74.00, and Group C (dual-modality) scored 75.33. Although the differences were statistically non-significant, this incremental pattern suggests potential benefits of dual-modality delivery for comprehension.

Group A exhibited the highest variability ( $SD=23.30$ ), with scores ranging from 20 to 100, compared to narrower ranges in Group B (40–100) and Group C (30–100). Notably, all groups included participants achieving perfect scores (100), indicating that each modality could support optimal comprehension under favorable conditions. Group C, however, combined the highest mean score with a relatively homogeneous distribution ( $SD=18.85$ ), suggesting enhanced stability in information processing.

### Cognitive effort

Participants' neurological responses to three distinct content delivery modes were investigated through analysis of brain activity patterns across participant groups. The topographical distribution of the raw EEG data was visualized using the EEGLab toolbox [61] in Matlab R2022a, as illustrated in Figs 3 and 4. The analysis revealed distinct patterns of neural activity across the three groups, with particular emphasis on power bands, which are strongly associated with human cognitive performance.

Fig 3 presents the neural oscillations associated with theta waves, which have been recognized as reliable markers of cognitive effort [49], for the three groups. An increase in theta wave power is typically linked to tasks that demand greater cognitive effort, impose strain on working memory [50], or result in mental fatigue [53]. Moreover, theta waves serve as indicators of mental resource allocation during the processing of multimodal (audio-visual) content [51,52].

**Table 2. Comprehension scores of the three groups.**

Group	Max	Min	Mode	Mean	SD
Group A	100	20	90	70.00	23.30
Group B	100	40	80	74.00	15.95
Group C	100	30	90	75.33	18.85

Notes: Group A: Automatic subtitles; Group B: SI; Group C: Automatic subtitles + SI.

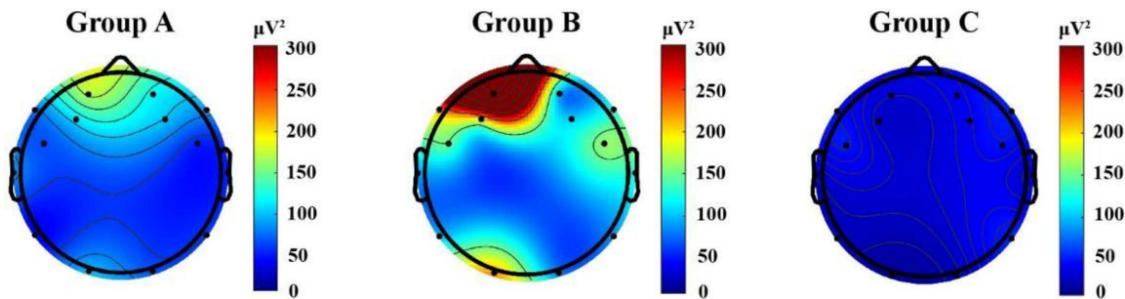
<https://doi.org/10.1371/journal.pone.0330692.t002>

As shown in Fig 3, Group B demonstrated the highest theta band activation in both the frontal region (AF3) and the occipital region (O7), followed by Group A, with lower theta band activation in the frontal region (AF3), and then Group C, with limited theta band activation.

**Viewing experience**

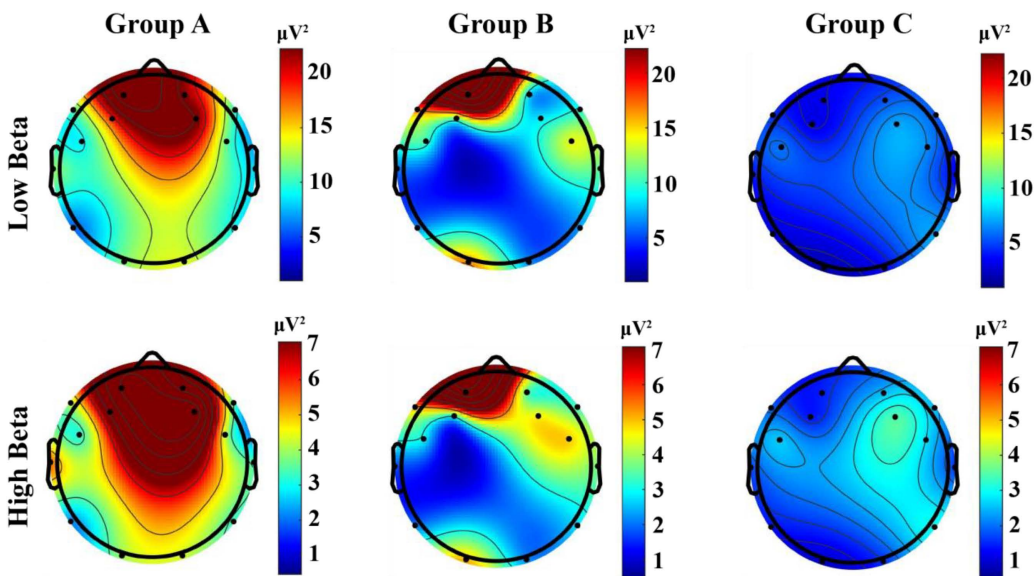
Regarding the viewing experience, Fig 4 illustrates the hierarchical pattern of beta waves associated with active cognitive processing, problem-solving, and alertness, providing insights into the participants' engagement levels and potential stress responses [60,62]. Elevated beta waves activity is generally correlated with excessive levels may indicate anxiety or stress [60,62]. Results show that Group A exhibited the highest beta band activation, indicating the highest stress level, followed by Group B, with Group C showing the lowest level of stress.

In addition to the raw EEG data, which were subsequently drawn as a topographic brain activity heatmap, the Emotiv Pro Suite software package was used to obtain stress values for the participants from the EEG devices while watching the 220-second conference video. We performed the non-parametric Kruskal-Wallis *H* test for the three groups, and the



**Fig 3. Topographic brain activity heatmap of EEG data in the theta band.**

<https://doi.org/10.1371/journal.pone.0330692.g003>



**Fig 4. Topographic brain activity heatmap of EEG data in the beta bands.**

<https://doi.org/10.1371/journal.pone.0330692.g004>

results are shown in Fig 5. There were significant differences in the pressure values of the three groups  $p < 0.01$ , demonstrating the accuracy and rationality of the discrepancies in the beta band.

## Discussion

### How do automatic subtitles compare to SI in facilitating content comprehension for viewers?

The experimental results revealed that the mode of delivery of translated information across the three groups did not significantly impact viewers' comprehension scores. This aligns with findings by Szarkowska et al. [19], who observed that the absence of audio input had only a modest effect on comprehension when subtitles were available. In addition, Liao et al. [63] found that when participants watched videos with subtitles but without accompanying audio, comprehension scores remained similar to those of participants who had access to both modalities. Other studies have further supported this observation, finding that even the non-syntactic segmentation of subtitles (i.e., splitting sentences at unnatural points) does not negatively impact viewers' comprehension [64,65].

While our statistical analysis revealed only minor differences in comprehension scores across experimental conditions, the group exposed to a combination of automatic subtitles and SI achieved the highest average comprehension scores. This finding highlights the potential of multimodal translation delivery to enhance participants' understanding [19]. By distributing information across auditory and visual channels, the dual-modality approach allows participants to process content more efficiently, as cognitive resources can be shared between modalities, reducing the strain on any single processing channel.

This distribution aligns with the cognitive theory of multimedia learning [66,67], which suggests that synchronizing corresponding visual and auditory material reduces cognitive load. Subtitles can clarify ambiguous or unfamiliar auditory input, while auditory content provides additional context or emotional nuance that subtitles alone may not fully convey [19]. The complementary nature of these modalities enhances understanding while minimizing the risk of cognitive overload, particularly in complex or fast-paced content. Consequently, multimodal delivery systems, such as the combination of automatic subtitles and SI, represent an effective and accessible strategy to support participants' comprehension, especially in multilingual or cognitively demanding environments.

### What differences in cognitive demands do automatic subtitles and SI impose on viewers?

The experimental findings yielded important insights into cognitive effort during conference content processing across the three experimental groups. Analysis of EEG topographic brain activity indicated significant variations in cognitive effort between the SI (listening) and automatic subtitles (reading) conditions, offering valuable theoretical and practical implications.

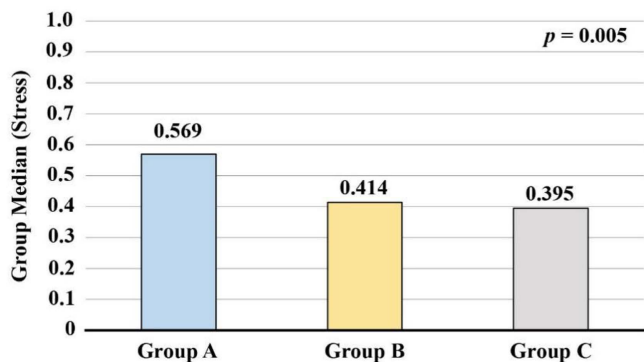


Fig 5. Non-parametric Kruskal-Wallis  $H$  test results (stress) for the three groups.

<https://doi.org/10.1371/journal.pone.0330692.g005>

The theta waves measurements demonstrated a hierarchical pattern of cognitive demand: Group B (SI only) exhibited the highest theta power, followed by Group A (automatic subtitles only), while Group C (combined SI and automatic subtitles) showed the lowest level. This finding is particularly relevant to conference interpreting, as conference attendees relying on translation services experience varying levels of cognitive load depending on the modality through which they receive translated information. The heightened cognitive effort in the SI-only condition suggests additional mental resources were needed for participants to comprehend by processing only auditory information while those listened to SI likely due to the mental effort required to process and comprehend auditory input in real time. In contrast, attendees using automatic subtitles only showed lower cognitive demand, as reading visual text may involve a less resource-intensive processing pathway. Most notably, attendees who had access to both SI and auto-subtitles, exhibited the lowest theta power, indicating reduced cognitive strain when provided with multimodal inputs. The minimal cognitive effort observed for Group C finds robust theoretical support from the bilingual dual-coding theory [68] that suggests memory performance improves when information is processed through multiple channels, in our case the visual and audio channels. These findings extend beyond simple cognitive load measurement, suggesting that the traditional reliance on audio-only SI may not fully optimize audience comprehension and engagement.

The comparison of Groups A and B reveals an interesting deviation from conventional assumptions that watching subtitle videos without corresponding audio would increase cognitive load as viewers use metacognitive strategies to adjust their visual processing to maintain comprehension [19,69]. Audio input is transient and sequential, while visual input is often static and simultaneous, allowing viewers to process information at their own pace and revisit it as needed. Our finding may reflect the experimental setting, which simulated a conference setting where attendees typically listen to speakers for various purposes. In this context, participants were motivated to understand the experimental material, whether through auto-subtitles, SI, or a combination of both, potentially requiring additional cognitive resources for comprehension and retention. Our findings suggest that the professional context significantly may influence how different modalities affect cognitive processing.

### How do automatic subtitles and SI influence viewers' overall viewing experience?

Beta wave analysis revealed distinct stress patterns across the three groups in relation to the viewing experience, providing insights on how different modes of translation delivery impact the overall viewing experience. Elevated beta activity is closely associated with increased mental workload and emotional tension, making it a reliable physiological marker for cognitive strain in audiovisual tasks [55]. Previous studies have employed beta wave analysis to assess user experiences in various applied contexts, such as game design and operational platform usability [46,47].

In this study, participants in Group A (automatic subtitles only) exhibited the highest beta wave levels, suggesting the greatest cognitive load, potentially diminishing subjective viewing comfort. This could be attributed to the demands of processing complex professional content solely through the visual modality. As Szarkowska et al. [19] observed, in the absence of auditory input in the first language (L1), viewers often adopt compensatory metacognitive strategies, such as reading subtitles more thoroughly, which can increase cognitive effort and stress.

Participants in Group B (SI only) showed lower beta wave levels compared to participants in Group A, suggesting a relatively smoother viewing experience. The real-time auditory input in SI likely facilitated reduced cognitive load by conveying prosodic features—such as tone, emotional emphasis, and pragmatic cues—that are inherently absent in text-based automatic subtitles. However, interpreting-based delivery is not without its own constraints. As prior studies have shown [70,71], the transient nature of spoken input places substantial demands on working memory, especially when the content is fast-paced or linguistically dense. Viewers must maintain high levels of attention, with no opportunity to pause, replay, or visually reinforce missed information.

Group C, which combined automatic subtitles and SI, demonstrated the lowest beta wave levels across all groups, suggesting the most cognitively relaxed and effective viewing experience. This dual-modality approach could balance

cognitive load by leveraging the complementary strengths of both modalities. Automatic subtitles serve as a visual reinforcement, offering viewers extended exposure to verbal content that can support decoding or fill in missed auditory information. Meanwhile, SI enriches the experience by providing expressive and contextual nuances that subtitles may fail to capture. The redundancy in information delivery across modalities allows for greater processing flexibility: when one channel is momentarily ineffective or ambiguous, the other can compensate [66]. This integration not only reduces stress but also enhances comprehension and information retention.

These findings align with the Cognitive Theory of Multimedia Learning [66,67], which posits that when auditory and visual information are temporally aligned, cognitive resources are more efficiently distributed across dual channels. Such integration reduces the likelihood of overload in a single modality and enhances overall processing fluency. While our results do not confirm statistical significance in comprehension performance, the physiological evidence from EEG beta activity strongly suggests that dual-modality delivery promotes a more comfortable and cognitively sustainable viewing experience in demanding multilingual environments.

## Practical implications

The findings of this study offer several practical implications for reducing cognitive load and enhancing the usability of audiovisual content, particularly in contexts demanding SI and real-time accessibility. These insights are particularly relevant for online and hybrid conferences, where addressing the diverse needs of international audiences is critical to ensuring both engagement and comprehension.

First, this study reinforces the critical role of subtitle design in modulating cognitive load. Subtitle systems should not only prioritize linguistic accuracy but also be designed for cognitive accessibility, especially for non-native or linguistically disadvantaged users. In practice, this means ensuring optimal segmentation, timing, and readability, which can alleviate processing burden during complex or rapid content delivery. In multilingual classrooms or international live broadcasts, AI-powered subtitle systems, especially those informed by cognitive feedback metrics such as EEG or user-reported load [72,73], can support more inclusive and cognitively sustainable communication.

Second, the superior performance of the combined modality (SI with automatic subtitles) highlights the value of multimodal delivery in supporting comprehension and reducing stress. In specialized or technical conference settings, where terminology density and speech rate are often high, subtitles act as a visual scaffold, while SI provides prosodic and emotional cues that enhance contextual understanding. This dual-modality approach distributes processing demands across auditory and visual channels, thereby reducing cognitive overload and improving the overall viewer experience.

Moreover, although comprehension scores did not differ significantly across conditions, the dual-modality group exhibited both lower stress levels and the highest mean performance. These findings suggest that scalable, technology-driven multimodal systems can enhance cognitive resilience and flexibility. In cost-sensitive scenarios or instances where SI is unavailable, auto-subtitling may serve as a provisional alternative, provided that accuracy and synchronization are maintained. With continued advancements in AI and user-centered design, such systems can be refined to meet the increasing demand for inclusive and efficient multilingual communication in professional and globalized settings.

## Conclusions

This study investigates the combined effects of auto-subtitling and SI on viewers' comprehension, cognitive effort, and overall viewing experience in a conference interpreting setting. The results showed that integrating automatic subtitles with SI reduced cognitive effort and improved the overall viewing experience compared to either modality alone, though comprehension differences were not statistically significant. EEG data revealed that participants in the automatic subtitle-only condition (Group A) exhibited the highest beta activity, followed by Group B (SI-only), while the dual-modality condition (Group C) showed the lowest levels of stress. Interestingly, despite the greater cognitive effort, participants in Group B

rated their viewing experience more positively than those in Group A, suggesting that auditory input may provide additional engagement and relaxation.

Despite these promising results, our study has several limitations. These include a relatively small participant group, a short speech duration, and a focus on a single language pair, which may limit the generalizability of the findings. Future research should address these limitations by incorporating larger and more diverse samples, testing longer and more complex materials, and evaluating dual-modality systems across a broader range of language pairs and cultural contexts. In terms of system development, future research can build on these neurocognitive insights to explore adaptive technologies that integrate real-time EEG monitoring with AI-powered subtitle systems. Drawing inspiration from frameworks such as SUMART [72] and ClinClip's EEG-informed modeling [73], future systems could dynamically adjust subtitle presentation parameters, such as speed, segmentation, and lexical complexity, based on ongoing cognitive feedback. Incorporating large language models (LLMs) with neurophysiological inputs, as shown in recent studies on personalized language learning [74,75], may further support real-time user-centered adaptation. Finally, future work should also consider extending accessibility beyond audio-visual modes by incorporating sign language interpreting.

## Supporting information

### S1 File. Questionnaire in EN.

(DOCX)

### S2 File. EEG based data.

(XLSX)

### S3 File. Comprehension data.

(XLSX)

## Author contributions

**Conceptualization:** Yanlin Li.

**Data curation:** Jiawen Diao.

**Formal analysis:** Yanlin Li.

**Investigation:** Jiawen Diao.

**Methodology:** Yanlin Li, Jiawen Diao.

**Project administration:** Jiawen Diao.

**Resources:** Jiawen Diao.

**Supervision:** Andrew K. F. Cheung.

**Validation:** Yanlin Li.

**Visualization:** Yanlin Li.

**Writing – original draft:** Yanlin Li.

**Writing – review & editing:** Andrew K. F. Cheung.

## References

1. Cintas JD. Subtitling: Theory, practice and research. In: *The Routledge handbook of translation studies*. Routledge; 2013. p. 273–87.
2. Cintas JD, Remael A. *Subtitling: Concepts and practices*. Routledge; 2020.
3. Basari A, Nugroho RA. The use of Aegisub in teaching audiovisual translation classes: a review on IT-based subtitling course. In: *Proceedings Education and Language International Conference*, vol. 1, no. 1. 2017.

4. Orrego-Carmona D. Audiovisual translation and audience reception. In: *The Routledge Handbook of Audiovisual Translation*. Routledge; 2018. p. 367–82. <https://doi.org/10.4324/9781315717166-23>
5. Tardel A. Effort in semi-automatized subtitling processes: speech recognition and experience during transcription. *J Audiovisual Translat*. 2020;3(2):79–102.
6. Lunin M, Minaeva L. Translated Subtitles Language Learning Method: A New Practical Approach to Teaching English. *Procedia Soc Behav Sci*. 2015;199:268–75. <https://doi.org/10.1016/j.sbspro.2015.07.516>
7. Vanderplank R. 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Lang Teach*. 2016;49(2):235–50. <https://doi.org/10.1017/s0261444813000207>
8. Jung H. Subtitles in South Korean entertainment television: Focused on a third-person effect. *J Entertain Media Stud*. 2016;2(1):78.
9. Li T, Chmiel A. Automatic subtitles increase accuracy and decrease cognitive load in simultaneous interpreting. *Interpreting*. 2024;26(2):253–81. <https://doi.org/10.1075/intp.00111.li>
10. Julaiti K, Cheung NDY, Cheung AKF, Huang JYJ. Number training in simultaneous interpreting: A corpus-assisted longitudinal study. *Interpret Soc*. 2025;5(1):80–96. <https://doi.org/10.1177/27523810241285542>
11. Cheung AKF. Listeners' perception of the quality of simultaneous interpreting and perceived dependence on simultaneous interpreting. *Interpreting*. 2022;24(1):38–58. <https://doi.org/10.1075/intp.00070.che>
12. Cheung AKF. Cognitive load in remote simultaneous interpreting: place name translation in two Mandarin variants. *Humanit Soc Sci Commun*. 2024;11(1):1238. <https://doi.org/10.1057/s41599-024-03767-y>
13. Garza TJ. Evaluating the Use of Captioned Video Materials in Advanced Foreign Language Learning. *Foreign Lang Annals*. 1991;24(3):239–58. <https://doi.org/10.1111/j.1944-9720.1991.tb00469.x>
14. Danan M. Captioning and Subtitling: Undervalued Language Learning Strategies. *Meta Journal des Traducteurs/Translators' Journal*. 2004;49(1):67–77. <https://doi.org/10.7202/009021ar>
15. Gant Guillory H. The Effects of Keyword Captions to Authentic French Video on Learner Comprehension. *CALICO J*. 1998;15(1–3):89–108. <https://doi.org/10.1558/cj.v15i1-3.89-108> PMID: 750440780
16. van der Zee T, Admiraal W, Paas F, Saab N, Giesbers B. Effects of Subtitles, Complexity, and Language Proficiency on Learning From Online Education Videos. *J Media Psychol*. 2017;29(1):18–30. <https://doi.org/10.1027/1864-1105/a000208>
17. Kruger JL, Doherty S, Soto-Sanfiel MT. Original language subtitles: their effects on the native and foreign viewer. *Sci J Media Educ*. 2017;25(50):23–32.
18. Liao S, Kruger J-L, Doherty S. The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *J Special Trans*. 2020;(33):70–98. <https://doi.org/10.26034/cm.jostrans.2020.549>
19. Szarkowska A, Ragni V, Szkriba S, Black S, Orrego-Carmona D, Kruger J-L. Watching subtitled videos with the sound off affects viewers' comprehension, cognitive load, immersion, enjoyment, and gaze patterns: A mixed-methods eye-tracking study. *PLoS One*. 2024;19(10):e0306251. <https://doi.org/10.1371/journal.pone.0306251> PMID: 39374196
20. Malakul S, Park I. The effects of using an auto-subtitle system in educational videos to facilitate learning for secondary school students: learning comprehension, cognitive load, and satisfaction. *Smart Learn Environ*. 2023;10(1):4. <https://doi.org/10.1186/s40561-023-00224-2> PMID: 40477865
21. Defrancq B, Fantinuoli C. Automatic speech recognition in the booth: Assessment of system performance, interpreters' performances and interactions in the context of numbers. *Target*. 2021;33(1):73–102.
22. Tammasrisawat P, Rangponsumrit N. The Use of ASR-CAI Tools and Their Impact on Interpreters' Performance During Simultaneous Interpretation. *New Voices Translat Stud*. 2023;28(2).
23. Baranowska K. Learning most with least effort: subtitles and cognitive load. *ELT J*. 2020;74(2):105–15. <https://doi.org/10.1093/elt/ccz060>
24. Díaz-Cintas J. The name and nature of subtitling. In: Bogucki Ł, Deckert M, editors. *The Palgrave handbook of audiovisual translation and media accessibility*. Cham: Springer International Publishing; 2020. p. 149–71.
25. O'Sullivan C, Cornu JF. History of audiovisual translation. *The Routledge handbook of audiovisual translation*. New York: Routledge; 2019.
26. Alharbi S, Alrazgan M, Alrashed A, Alnomasi T, Almojel R, Alharbi R, et al. Automatic Speech Recognition: Systematic Literature Review. *IEEE Access*. 2021;9:131858–76. <https://doi.org/10.1109/access.2021.3112535>
27. Cintas JD, Remael A. *Audiovisual translation: Subtitling*. London: Routledge; 2007.
28. Davis KH, Biddulph R, Balashek S. Automatic Recognition of Spoken Digits. *The J Acoust Soc Am*. 1952;24(6):637–42. <https://doi.org/10.1121/1.1906946>
29. Zhang Q, Lu H, Sak H, Tripathi A, McDermott E, Koo S, et al. Transformer transducer: A streamable speech recognition model with transformer encoders and rnn-t loss. In: *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020. p. 4–8.
30. Fujimoto M, Kawai H. One-pass single-channel noisy speech recognition using a combination of noisy and enhanced features. In: *Interspeech Conference, Int Speech Commun A*, editors. Graz, Austria: 2019.
31. Negi S, Mitra R. Native language subtitling of educational videos: A multimodal analysis with eye tracking, EEG and self-reports. *Brit J Educational Tech*. 2022;53(6):1793–816. <https://doi.org/10.1111/bjet.13214>

32. d'Ydewalle G, Praet C, Verfaillie K, Rensbergen JV. Watching Subtitled Television. *Commun Res.* 1991;18(5):650–66. <https://doi.org/10.1177/009365091018005005>
33. Gorman BM, Crabb M, Armstrong M. Adaptive subtitles: Preferences and trade-offs in real-time media adaptation. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama, Japan: Association for Computing Machinery; 2021. p. 733.
34. Becerra H, Ragano A, Debnath D, Ullah A, Lucas CR, Walsh M, et al. Dialogue understandability: Why are we streaming movies with subtitles? PMID: INSPEC:25017235
35. Böser U, Deane-Cox S. *Translation, interpreting and technological change: Innovations in research, practice and training*. London: Bloomsbury Publishing; 2024.
36. González ER, Braun S, Davitti E, Korybski T. *The use of automatic speech recognition in cloud-based remote simultaneous interpreting (Doctoral dissertation)*. University of Surrey; 2024.
37. Folkman S. *Stress: Appraisal and coping*. Encyclopedia of behavioral medicine. Cham: Springer International Publishing; 2020.
38. Thoen A, Steyaert J, Alaerts K, Evers K, Van Damme T. A Systematic Review of Self-Reported Stress Questionnaires in People on the Autism Spectrum. *Rev J Autism Dev Disord.* 2023;10(2):295–318. <https://doi.org/10.1007/s40489-021-00293-4> PMID: 34603935
39. Westbrook A, Braver TS. Cognitive effort: A neuroeconomic approach. *Cogn Affect Behav Neurosci.* 2015;15(2):395–415. <https://doi.org/10.3758/s13415-015-0334-y> PMID: 25673005
40. Chen S, Kruger JL. The effectiveness of computer-assisted interpreting: A preliminary study based on English-Chinese consecutive interpreting. *Translat Interpret Stud.* 2023;18(3):399–420.
41. Seeber KG. Multimodal processing in simultaneous interpreting. *The handbook of translation and cognition*. 2017.
42. Das Chakladar D, Roy PP. Cognitive workload estimation using physiological measures: a review. *Cogn Neurodyn.* 2024;18(4):1445–65. <https://doi.org/10.1007/s11571-023-10051-3> PMID: 39104683
43. Aghajani H, Garbey M, Omurtag A. Measuring Mental Workload with EEG+fNIRS. *Front Hum Neurosci.* 2017;11:359. <https://doi.org/10.3389/fnhum.2017.00359> PMID: 28769775
44. Cano S, Araujo N, Guzman C, Rusu C, Albiol-Perez S. Low-Cost Assessment of User eXperience Through EEG Signals. *IEEE Access.* 2020;8:158475–87. <https://doi.org/10.1109/access.2020.3017685>
45. Ragni V. Physiology-based research in second language acquisition: New evidence for didactic audiovisual translation. In: *Empirical studies in didactic audiovisual translation*. Routledge. 2025. p. 7–25.
46. Li Y, Po Tsang Y, Lee CKM, Han S. Integrating neurophysiological sensing and group-based multi-criteria decision-making for fourth-party logistics platform selection. *Adv Eng Info.* 2025;64:102968. <https://doi.org/10.1016/j.aei.2024.102968>
47. Paranthaman PK, Bajaj N, Solovey N, Jennings D. Comparative evaluation of the eeg performance metrics and player ratings on the virtual reality games. In: *2021 IEEE Conference on Games (CoG)*. 2021. p. 17–20.
48. Castro-Meneses LJ, Kruger J-L, Doherty S. Validating theta power as an objective measure of cognitive load in educational video. *Education Tech Research Dev.* 2020;68(1):181–202. <https://doi.org/10.1007/s11423-019-09681-4>
49. Puma S, Matton N, Paubel P-V, Raufaste É, El-Yagoubi R. Using theta and alpha band power to assess cognitive workload in multitasking environments. *Int J Psychophysiol.* 2018;123:111–20. <https://doi.org/10.1016/j.jpsycho.2017.10.004> PMID: 29017780
50. Klimesch W, Schack B, Sauseng P. The functional significance of theta and upper alpha oscillations. *Exp Psychol.* 2005;52(2):99–108. <https://doi.org/10.1027/1618-3169.52.2.99> PMID: 15850157
51. Boos M, Kobi M, Elmer S, Jäncke L. The influence of experience on cognitive load during simultaneous interpretation. *Brain Lang.* 2022;234:105185. <https://doi.org/10.1016/j.bandl.2022.105185> PMID: 36130466
52. Grabner RH, Brunner C, Leeb R, Neuper C, Pfurtscheller G. Event-related EEG theta and alpha band oscillatory responses during language translation. *Brain Res Bull.* 2007;72(1):57–65. <https://doi.org/10.1016/j.brainresbull.2007.01.001> PMID: 17303508
53. Tran Y, Craig A, Craig R, Chai R, Nguyen H. The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. *Psychophysiology.* 2020;57(5):e13554. <https://doi.org/10.1111/psyp.13554> PMID: 32108954
54. Isabella SL, Cheyne JA, Cheyne D. Inhibitory Control in the Absence of Awareness: Interactions Between Frontal and Motor Cortex Oscillations Mediate Implicitly Learned Responses. *Front Hum Neurosci.* 2021;15:786035. <https://doi.org/10.3389/fnhum.2021.786035> PMID: 35002659
55. Lazarus RS, Folkman S. *Stress, appraisal, and coping*. Springer Publishing Company; 1984.
56. Kurz I. A look into the 'black box'—EEG probability mapping during mental simultaneous interpreting. *Translation Stud.* 1994;2:199.
57. Kurz I. *Watching the brain at work—an exploratory study of EEG changes during Simultaneous Interpreting (SI)*. 1995.
58. Koshkin R, Shtyrov Y, Myachykov A, Ossadtschi A. Testing the efforts model of simultaneous interpreting: An ERP study. *PLoS One.* 2018;13(10):e0206129. <https://doi.org/10.1371/journal.pone.0206129> PMID: 30356337
59. Moser-Mercer B. Simultaneous interpreting: Cognitive potential and limitations. *Interpreting.* 2000;5(2):83–94.
60. Yagura H, Tanaka H, Kinoshita T, Watanabe H, Motomura S, Sudoh K, et al. Selective Attention Measurement of Experienced Simultaneous Interpreters Using EEG Phase-Locked Response. *Front Hum Neurosci.* 2021;15:581525. <https://doi.org/10.3389/fnhum.2021.581525> PMID: 34163336

61. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. 2004;134(1):9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009> PMID: [15102499](https://pubmed.ncbi.nlm.nih.gov/15102499/)
62. Palva S, Palva JM. New vistas for alpha-frequency band oscillations. *Trends Neurosci*. 2007;30(4):150–8. <https://doi.org/10.1016/j.tins.2007.02.001> PMID: [17307258](https://pubmed.ncbi.nlm.nih.gov/17307258/)
63. Liao S, Yu L, Kruger J-L, Reichle ED. The impact of audio on the reading of intralingual versus interlingual subtitles: Evidence from eye movements. *Appl Psychol*. 2021;43(1):237–69. <https://doi.org/10.1017/s0142716421000527>
64. Gerber-Morón O, Szarkowska A, Woll B. The impact of text segmentation on subtitle reading. *J Eye Mov Res*. 2018;11(4):10.16910/jemr.11.4.2. <https://doi.org/10.16910/jemr.11.4.2> PMID: [33828704](https://pubmed.ncbi.nlm.nih.gov/33828704/)
65. Perego E, Del Missier F, Porta M, Mosconi M. The Cognitive Effectiveness of Subtitle Processing. *Media Psychol*. 2010;13(3):243–72. <https://doi.org/10.1080/15213269.2010.502873>
66. Mayer RE, Moreno R. Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Education Psychol*. 2003;38(1):43–52. [https://doi.org/10.1207/s15326985ep3801\\_6](https://doi.org/10.1207/s15326985ep3801_6)
67. Mayer RE. Cognitive Theory of Multimedia Learning. *Cambridge Handbook Multimedia Learn*. 2005;41(1):31–48.
68. Paivio A, Desrochers A. A dual-coding approach to bilingual memory. *Canadian J Psychol/ Revue canadienne de psychologie*. 1980;34(4):388–99. <https://doi.org/10.1037/h0081101>
69. Schnotz W. An integrated model of text and picture comprehension. In: *The Cambridge handbook of multimedia learning*. West Nyack: Cambridge University Press; 2005.
70. Justice LM, Lomax R, O'Connell A, Pentimonti J, Petrill SA, Piasta SB, et al. Are working memory and behavioral attention equally important for both reading and listening comprehension? A developmental comparison. *Read Writing*. 2018;31(7):1449–77.
71. Mayer RE. *The Cambridge Handbook of Multimedia Learning*. 2nd ed. Cambridge Handbooks in Psychology. West Nyack: Cambridge University Press; 2014.
72. Nishida N, Rekimoto J. SUMART: SUMmARizing Translation from Wordy to Concise Expression. In: *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE; 2024. p. 613–5.
73. Sun G. ClinClip: a Multimodal Language Pre-training model integrating EEG data for enhanced English medical listening assessment. *Front Neurosci*. 2025;18:1493163. <https://doi.org/10.3389/fnins.2024.1493163> PMID: [39850622](https://pubmed.ncbi.nlm.nih.gov/39850622/)
74. Chen Y, Wang W, Yan S, Wang Y, Zheng X, Lv C. Application of electroencephalography sensors and artificial intelligence in automated language teaching. *Sensors*. 2024;24(21):6969.
75. Huang Y, Li D, Cheung AKF. Evaluating the linguistic complexity of machine translation and LLMs for EFL/ESL applications: An entropy weight method. *Res Method Appl Linguist*. 2025;4(3):100229. <https://doi.org/10.1016/j.rmal.2025.100229>