Parallèles 37(1) April 2025

Gesture and cognitive load in simultaneous interpreting: A pilot study

Yuetao Ren

University of International Business and Economics (China)

Jianhua Wang

Renmin University of China

Abstract

This paper explores the relationship between gesture and cognitive load in simultaneous interpreting (SI). To this end, we set up a remote interpreting setting for data collection. Thirteen master's student interpreters participated in two SI tasks, one in a video condition and the other in an audio condition. We analyzed their gestural behaviors and disfluency patterns, as well as the correlation and temporal relation between gestures and disfluencies. We found that interpreters gestured more in tasks with a higher cognitive load (audio interpreting), although the differences in disfluency rate and gesture rate between the two conditions were not significant. Even though the correlation between gesture and cognitive load was not significant, all the gestures in the study were produced parallel with or adjacent to processing difficulties. We conclude that gestures could be the embodied manifestation of the cognitive processes of SI and of the 'exported load'. Furthermore, the function of each gesture type varies under cognitive load. Silent gestures (beats and metaphorics) may reflect the interpreter's use of strategies, while the production of semantically related gestures (deictics and iconics) may be influenced by cognitive load. The results contribute to the understanding of SI as an embodied, multimodal cognitive activity.

Keywords

Simultaneous interpreting, gesture, cognitive load, disfluency, multimodality

 DOI:10.17462/para.2025.01.04

1. Introduction

Interpreting is generally acknowledged as a cognitively demanding activity (Chen, 2017). Simultaneous interpreting (SI), one of the major working modes of interpreting in conference settings, is considered the most complex language task (Christoffels & De Groot, 2005). SI is often studied from a cognitive perspective, focusing on its multitasking nature and high cognitive demand.

In recent years, the idea that cognition is embodied, embedded, extended, enactive, and affective has been adopted in the cognitive study of translation and interpreting (Muñoz Martín, 2016). In this paper, we adopt the embodied approach to cognition, which holds that cognitive processes are deeply rooted in the body's interactions with the world (Wilson, 2002). Embodiment explains thinking and speaking as perceptually and motorically based (Barsalou, 2008). The representation and manipulation of information is accomplished via the simulation of sensorimotor processes, and these simulation mechanisms may give rise to both speech and gesture (Hostetter & Alibali, 2008). When performing a complex language task like SI, the interpreter employs a wide range of embodied modalities like gesture, gaze, and posture. Among these modalities, gestures are intimately connected with language and have been shown to perform self-oriented functions, facilitating the cognitive processes of thinking and speaking (Kita *et al*., 2017).

Studies on the multimodal behavior of interpreters have revealed that interpreters do gesture during SI, even when they are 'invisible' in the booth (Adam & Castro, 2013; Cienki, this volume; Cienki & Iriskhanova, 2020; Martín de León & Fernández Santana, 2021; Martín de León & Zagar Galvão, this volume; Zagar Galvão, 2020). However, a systematic study of the embodied, multimodal cognitive processes of SI is still lacking. Gestures are outward manifestations of the cognitive processes that govern thinking and speaking. They open a "window onto the mind" (McNeill, 1992, p. 268). As such, gestures may provide visual clues for the study of the cognitive processes of SI. This study is a pilot for such research. We include the gestural behaviors of simultaneous interpreters in our analysis while focusing on one of the most distinctive aspects of SI, i.e. cognitive load. Before introducing the methodology, we will review the theoretical aspects of cognitive load in SI, discuss the role of gestures in thinking and speaking, and comment on the existing studies of gesture and cognitive processes of SI.

1.1. Cognitive load in simultaneous interpreting

Following Seeber (2013, p. 19), we define cognitive load as "the amount of capacity the performance of a cognitive task occupies in an inherently capacity-limited system." This definition is based on the assumption that working memory is a capacity-limited system (Seeber, 2011). In other words, the number of operations the human brain can carry out and the amount of information it can maintain for processing at a given time is limited. There are two models that offer explicit accounts for cognitive load in interpreting: Effort Models (Gile, 1999, 2008) and the Cognitive Load Model (Seeber, 2011; Seeber & Kerzel, 2012).

Effort Models are a set of models that account for operational constraints in interpreting (Gile, 1999). Interpreting is conceptualized as a set of multiple cognitive operations which can be grouped into certain 'Efforts', such as listening and analysis (L), speech production (P), and short-term memory (M) (Gile, 2008). Effort Models assume a single pool of resources for all efforts. It is further assumed that different efforts may compete for the total available processing capacity ('Competition Hypothesis') and interpreters tend to work close to saturation ('Tightrope Hypothesis') (Gile, 1999). Problem triggers, such as numbers, lists, and proper names, are associated with the increase of processing capacity requirements, which may exceed the total amount of capacity available or cause energy management problems,

resulting in errors, omissions, and/or reduced quality in the performance. Furthermore, the load might be carried over into downstream segments ('exported load') (Gile, 2008). Problem triggers do not necessarily lead to actual problems in the corresponding segments but may affect other segments, which are at a distance and are not difficult to render in and of themselves.

The Cognitive Load Model accounts for the effect of different combinations of sub-tasks on overall cognitive demands (Seeber, 2011). The load is not generated from the competition of resources, but rather from the interference between sub-tasks. The Cognitive Load Model distinguishes different processing codes (verbal-spatial) and modalities (auditory-visual). The former refers to the different systems of working memory, and the latter refers to the sensory modalities of input and response. According to this model, tasks of the common structures interfere with each other more strongly than those of different structures. In other words, processing codes and modalities have an effect on task interference. Discrete spatial and verbal tasks are time-shared more efficiently than two spatial or two verbal tasks, and intra-modal processes interfere with each other more than inter-modal processes (Seeber, 2007).

Effort Models and the Cognitive Load Model share some common ground. As Seeber (2011) argues, the two models aim to account for the cognitive demands inherent to interpreting. They both see interpreting as multitasking, which is comprised of a set of sub-tasks. Furthermore, Seeber and Kerzel (2012) present empirical evidence that corroborates Gile's (2008) idea of exported load, as the relative maximum local cognitive load occurs at the end of the sentence.

1.2. The role of gesture in thinking and speaking

1.2.1 Classifications of gesture

Co-speech gestures are taken here to be spontaneous speech accompaniments that are made with fingers, hands, and arms (McNeill, 2005). Gestures can be classified into several subtypes. McNeill (1992) distinguished four types in terms of the forms and functions of gestures: iconics, metaphorics, beats, and deictics. Iconic gestures represent concrete concepts by depicting the shape, size, or contour of the referent. Metaphoric gestures are similar to iconics in form but are associated with abstract concepts. Beats are biphasic movements of the finger or hand, which can serve an emphatic function. Deictic gestures are pointing movements that refer to an entity or a space by extending the finger, hand, or arm.

Kendon (2004) outlined three main ways in which gestures contribute to the meanings of utterances, including via referential, pragmatic, and interactional functions. Referential gestures contribute to the propositional meanings of the utterances. Pragmatic gestures contribute to the acts accomplished by utterances, such as indicating the speaker's attitude, providing an interpretative framework, or making manifest the speech act. Interactional gestures are used to regulate interactions. The referential and pragmatic functions of gestures are not mutually exclusive. For example, pointing gestures mainly serve referential functions, but the different hand shapes used in pointing may have pragmatic functions.

In this paper, we adopt McNeill's (1992) categorization of gestures, which is based on a psycholinguistic perspective, and gestures were perceived as a "window onto the mind" (p. 268).

1.2.2 Gesture-speech integration

Gesture and speech form an integrated system during language production (McNeill, 1992, 2005). The minimal idea unit is a 'growth point' (GP), which consists of both imagery and linguistic content and can develop into a full utterance with a gesture (McNeill, 1992, 2005). In other words, gesture and language share the same computational stage (McNeill, 1985), and

the product of this stage is a concept that can be packaged and expressed both in speech and in gesture.

Gesture either synchronizes with a parallel linguistic unit or comes *before* the linguistic expression, suggesting that gesture can reveal the moment at which the speaker formulates a concept (McNeill, 1985). McNeill distinguished two kinds of 'gestural anticipation' (p. 361). First, during uninterrupted speech, semantic computation takes place and is expressed in gesture, while the corresponding linguistic expression for the same concept may be delayed; it comes *after* the linguistic segment that goes with the gesture. Such gestures include iconic and deictic gestures; they refer to the content of the utterance and perform a referential function (Kendon, 2004). Second, during silence, where speech comes to a halt, there is no semantic computation taking place. Beat or metaphoric gestures may be produced during silence. Such 'silent gestures' are part of speaking; they provide a metalinguistic commentary on the process of speaking (McNeill, 1985, p. 354), fulfilling a pragmatic function (Kendon, 2004).

Gesture is involved in cognitive processes by activating, manipulating, packaging, and exploring information for thinking and speaking (Kita *et al*., 2017). Gestures have been argued to be generated from the same process that generates practical actions (Hostetter & Alibali, 2008). Thus, gestures can influence thoughts about both spatio-motoric information based on bodily experiences, and about abstract information, via the metaphorical use of spatio-motoric information. However, gestures are different from practical actions in that they are schematic representations. Focusing on essentials rather than details, such representations can be processed efficiently and are flexible and modifiable (Kita *et al*., 2017). As such, gestures play a facilitative role in the cognitive processes of thinking and speaking.

1.3. Gestures and the cognitive processes of interpreting

Recently, interpreting has been conceived as a multilingual and multimodal embodied cognitive activity (Cienki & Iriskhanova, 2020; Martín de León & Fernández Santana, 2021; Stachowiak-Szymczak, 2019). The input interpreters receive is inherently multimodal, including both auditory and visual information. When producing the output, interpreters employ a range of embodied modalities and resources. One of the most essential and inherent resources for interpreters is gesture. In conference interpreting, researchers have focused on the cognitive aspects of the interpreter's gestures.

Adam and Castro (2013) investigated the form and function of beat gestures produced by student interpreters during SI. Results showed that beats were the most prevalent (84.7%) gestures produced by interpreters, while 18.41% of all the gestures appeared at moments of hesitation. When gestures appeared during pauses, interpreters were having comprehension problems or engaging in word searches. These gestures could have been used as an unconscious strategy by participants. When gestures accompanied self-corrections, they usually went with the corrected version of the utterance for emphasis.

Stachowiak-Szymczak (2019) also focused on interpreters' beat gestures in both simultaneous and consecutive interpreting (CI). Using numbers and lists as problem triggers, the study tested different levels of cognitive effort and correlated them with the gestural behaviors of the interpreter. Results showed that cognitive effort was reflected in gesture numbers. Both professional and student interpreters produced more gestures when interpreting lists and numbers compared to interpreting narratives. Beat gestures could have been produced by participants to deal with local cognitive effort in interpreting, especially for reducing the load related to processing lists.

Martín de León and Fernández Santana (2021) analyzed the interpreting process of one professional simultaneous interpreter. The study revealed the different roles of the interpreter's

gestures in the interpreting process. Using referential gestures may help the interpreter with source language (SL) comprehension, while producing pragmatic gestures could lend support to her target language (TL) production.

Taken together, results of the above studies show that gestures employed by interpreters are related to the cognitive processes of interpreting, especially at moments when cognitive demands are high. However, systematic analysis of gestures and cognitive processes is lacking. Stachowiak-Szymczak (2019) focused on only one type of gesture, while Martín de León and Fernández Santana (2021) studied only one interpreter. In this study, we try to expand the scope of previous studies by including all the co-speech gestures produced by eleven interpreters.

This study aims to explore the relationship between the interpreter's gestural behavior and cognitive load in simultaneous interpreting. It was guided by the following research questions:

1) Is there a correlation between the interpreter's gesture and cognitive load in simultaneous interpreting?

2) If so, what functions do gestures of different types play under cognitive load?

For the first question, we expect that more gestures are produced under high cognitive load. Following Gumul (2021), we used disfluencies as indicators of cognitive load. The assumption is that disfluencies are evidence of a decrease in interpreting quality, which is likely to be associated with an increase in cognitive load (Chen, 2017, p. 647). In other words, we expect that gestures are produced more with disfluent speech than with fluent speech (Cienki, this volume).

For the second question, we formulated two hypotheses about the functions of gestures of different types under cognitive load. Based on McNeill's (1985) notion of gestural anticipation, we categorized gestural behaviors under cognitive load into two kinds. Silent gestures are associated with processing difficulty, as they often arise during speech breakdown. In contrast, gestures produced alongside uninterrupted speech might not be accompanied by processing difficulties. Following McNeill's (1992) categorization of gesture types, we expect that all gesture types, including iconics, metaphorics, beats, and deictics, are produced in the interpreting process; but only beat and metaphoric gestures are produced with processing difficulty (silent gesture). Moreover, we expect that deictic and iconic gestures are produced before the production of their linguistic counterparts (gestural anticipation), without processing difficulty.

2. Methodology

2.1. Participants

Thirteen Chinese students in a Master of Translation and Interpreting program (MTI) (12 females and 1 male) participated in the study on a voluntary basis. The average age was 23.9 years (*SD* = 1.18 years, range 22 – 26 years). They had the same language combination, with Mandarin Chinese as L1 and English as L2. They all have passed the English language test TEM-8[1], with an average score of 75 (*SD* = 3.25, range 71 – 81). None of them had worked as a professional interpreter before. By the time of the experiment, they had received CI training for two semesters and SI training for one semester. Two participants (both females) did not perform any co-speech gestures in any of the interpreting tasks, but they did use such gestures in other speech production tasks in the experiment. Since our focus was co-speech gestures in SI, we excluded them from the data analysis for this study. The final *N* here was 11 (10 females and 1 male).

[1] TEM-8, which stands for Test for English Majors Band 8, is a Chinese equivalent of IELTS. Its full mark is 100.

All participants were tested to be right-handed using the Edinburgh Inventory Handedness Questionnaire[2]. They were told that the whole experiment would be videotaped and they signed a written consent before the experiment. After the experiment, they were given another consent form which concerned their willingness to have their video images used in academic reports. To protect the anonymity and confidentiality of participants, they were informed that their faces would be blocked when using their images. All participants agreed to participate in the experiment and to have their images used anonymously. They were given 100 RMB compensation for their time and efforts in participating in the experiment.[3]

2.2. Materials

We selected two speech videos of different topics from the same speaker. Speech A was adapted from an extended talk[4]. We selected the first five minutes and a half starting from the beginning, in which the speaker illustrated her first point of the whole talk. Speech B is a complete TED talk[5].

We calculated the text complexity of the spoken texts using the Flesch Kincaid Readability Index[6], considering the reading ease (A: 72.7, B: 70.2) and the percentage of complex words (A: 10.67%, B: 10.20%). The Flesch Reading Ease score ranges from 1 to 100. The two texts correspond to a school grade level 8 (ages 12 to 14) and should be fairly easy for the average adult to read. Besides, the percentage of complex words in the two texts is close to one another, reflecting a similar complexity of the two texts.

In terms of speed, the original rate of speaking in speech A and B were similar (145.41 wpm vs. 145.19 wpm respectively). For SI, the speed of the source speech is a potential 'problem trigger' (Gile, 2008). Since participants were not professional interpreters and had only mastered basic skills for SI, fast speed would pose a challenge for them. We made some minor adjustments to slow down the original speed.

We randomly selected speech A and converted it into an audio file, while the manipulated speed remained unchanged. Thus, we produced two source speech conditions, namely one as a video and one as an audio-only recording. For the video speech (speech B), participants would hear and only see the speaker's image in the video, without captions, PPT slides, and other images. The rationale for having two interpreting conditions is the potential relevance of codes and modalities in cognitive load (Seeber, 2007, 2011). The comparison between the two conditions could unveil such potential effects.

The word count, final length, and speed of the materials, as well as corresponding interpreting conditions, are shown in Table 1.

	Source speech	
	Speech A	**Speech B**
Word count	807	861
Final length	6'10''	6'35''
Final speed	130.79 wpm	130.85 wpm
Interpreting condition	Audio	Video

Table 1. Summary of information about the materials for the experiment

[2] Available at http://www.brainmapping.org/shared/Edinburgh.php#

[3] The experiment was approved by the research ethics committee of Renmin University.

[4] https://www.youtube.com/watch?v=tBRjmjsbrcE

[5] https://www.ted.com/talks/angela_lee_duckworth_grit_the_power_of_passion_and_perseverance

[6] https://www.webfx.com/tools/read-able/

2.3. Procedure

We set up a simulated remote interpreting setting via Tencent Meeting[7], a Chinese equivalent of ZOOM, for data collection. Participants and the experimenter were in two separate rooms (partly as a result of Covid restrictions). They were connected via Tencent Meeting on two laptops. The experimenter played the speech files on his laptop, while the screen and sound were shared via Tencent Meeting and simultaneously displayed on the participant's laptop, which was placed on the right front of them. For the audio speech (speech A), participants would see a blank screen. They could not see the experimenter or other audience in either of the interpreting tasks. The whole process of the experiment was filmed by two cameras simultaneously from different perspectives. This design was intended to ensure that all hand movements, including finger movements, could be captured without obstruction.

In the warm-up part of the experiment, participants first talked freely in Chinese about their English learning experience. In this part, participants could get familiar with the remote interpreting condition, and the experimenter could check the state of the apparatus and internet connection. Then, participants did simultaneous interpreting in two different conditions. The order of the two sessions was randomized per participant. After each interpreting task, participants conducted a cued retrospection using their own videos as stimuli, in which they made spoken commentaries on their interpreting process. Participants could take a break of two to three minutes after the retrospection. Data were originally collected for the first author's PhD thesis from May to October 2022. Aside from the two SI tasks, participants also performed two CI tasks using different materials, each followed by a retrospection. The order of the four interpreting tasks was randomized for each participant. After interpreting, we conducted a semi-structured interview on the use of gestures in communication. For this study, we only focused on the two SI tasks. The CI data will be used for another study and are not presented here.

Twenty-four hours before the experiment, we gave participants some term lists, including proper names and unfamiliar jargon, with corresponding Chinese equivalents. Participants could preview the terms as preparation for the interpreting tasks. Before each interpreting task, they could review the terms for that specific task, but they could not refer to them during the process of interpreting.

We did not inform participants of the actual purpose before the experiment, because knowing this might potentially affect the spontaneity of their gestural behaviors. All they knew was that it would involve a remote interpreting experiment. It was only at the end of the experiment, during the debriefing, that participants got to know that their gestures were being studied. The whole session took about 100 minutes for each participant.

2.4. Data analysis

We focused on two types of product data for the present study, including gestures and disfluencies. We used the ELAN software (version 6.2) (Sloetjes & Wittenburg, 2008) for the annotation and analysis of data.

For disfluencies, we included the following sub-types: silent pause, filled pause, and false start. According to Han and An (2020), we chose 0.5s as the threshold for identifying silent pauses. Following Bóna and Bakti (2020), we defined a filled pause as a sound or syllable that does not contribute to the meaning of the sentence, like *uh* or *um* in English and '嗯' (*enn*) or '呃' (*err*) in Chinese. In this study, we used 'false start' as an umbrella term, which included restarts and truncation (Cienki & Iriskhanova, 2020), partial or whole word repetition, revision, broken words, and prolonged sounds (Bóna & Bakti, 2020).

[7] https://meeting.tencent.com/

For gestures, we focused on co-speech gestures, i.e. hand movements that had an intimate relationship with the co-occurring speech (McNeill, 2005). Adaptors, non-speech-motivated movements like touching one's body or manipulating external objects (Litvinenko *et al*., 2018, p. 7), were not included in the analysis. The number of gestures was counted based on the number of gesture phrases, which consist of a preparation phase and a stroke phase (Kendon, 2004). In the case of multiple strokes in one gesture phrase, we calculated each stroke as one gesture.

We then analyzed the distribution of gestures between fluent and disfluent speech. Following Gile (2008), we used sentences as the unit of analysis. Fluent speech was characterized by the absence of any interruptions or disruptions within the sentence boundaries. Thus, gestures accompanied by fluent speech only refer to those gestures that were embedded in a complete fluent sentence, where no disfluencies emerged within the sentence boundaries. We labeled this group as "G + Fluency".

Disfluent speech was defined as a sentence with disfluencies occurring in its boundaries. Gestures parallelling disfluent speech were divided into two groups. In some cases, gestures were embedded in a disfluent sentence but were not parallel with disfluencies. That is to say, gestures and disfluencies co-occurred at different places within a sentence. They were adjacent to each other, as they appeared in the same sentence. We labeled this group as "G + Quasi-disfluency". In other cases, gestures parallelled disfluencies per se. We labeled them as "G + Disfluency". We extended the boundaries of disfluencies a little bit by including the first phoneme or word that immediately preceded or followed the disfluency. This makes sense because the boundaries of gesture and disfluency may not completely overlap with each other, and there could be a short time lag (usually a few milliseconds) between their boundaries. For example, a gesture's onset and preparation phase could be performed within the pause, with the subsequent stroke phase overlapping with the first phoneme after the pause. Gestures and disfluencies partially overlapped with each other. Hence, we also included in this group gestures whose strokes were located on the first phoneme or word that immediately preceded or followed the disfluency.

The "G + Disfluency" group deserves more attention, for these were cases where gestures could be potentially related to the fluctuation of cognitive load. We gave a more detailed account of the temporal relation between gestures and disfluencies for this group. Given that we included the first phoneme or word before and after disfluency into its boundaries, three categories naturally emerged. We used "Pre-disfluency G" to refer to gestures whose stroke co-occurred with the first phoneme or word *before* disfluency. 'Peri-disfluency G" referred to gestures whose stroke was *within* disfluency. "Post-disfluency G" referred to gestures whose stroke was produced with the first phoneme or word *after* disfluency.

In some cases, a gesture may consist of multiple strokes. Some strokes were performed within a disfluency, while other strokes were performed with the following fluent word(s). Strokes that accompanied the fluent word were designated as 'Post-disfluency G', while only those strokes that overlapped with disfluencies per se were classified into the 'Peri-disfluency G' category.

Gesture annotations were conducted following the annotation procedures developed by Litvinenko and colleagues (Litvinenko *et al*., 2018). We calculated gesture rate and disfluency rate, which referred to the total number of gestures and the total number of disfluencies divided by the duration of the interpreting product, respectively. We used minutes as the unit of time. Statistical analyses were conducted in SPSS 26.0.

The coding of gestures and disfluencies was mainly conducted by the first author twice, with a time interval of one month. Intra-coder agreement was 83.9%, indicating a high degree of agreement. After that, the second author randomly coded 5% of the data, following the same procedures. Inter-coder agreement was 78.6%, reflecting a strong level of consistency. The two authors also discussed and resolved inconsistencies.

3. Results

3.1. Disfluency patterns in SI tasks

Altogether, we have identified 3,245 disfluencies in the SI dataset, including 2,262 silent pauses longer than 0.5 seconds (69.7%), 361 filled pauses (11.1%), and 622 false starts (19.2%). Filled pauses are usually comprised of a filler word, like *um* or *uh*, with silent pauses occurring before and/or after the filler word. In our analysis, if the silent part of the filled pause exceeds the threshold of 0.5 seconds, it is coded as a separate silent pause.

Disfluency rate refers to the number of disfluencies per minute (dpm). The mean disfluency rate was 20.03 dpm (*SD* = 3.71 dpm, range 14.05 – 30.97 dpm). Using disfluencies as indicators of cognitive load, we compared the cognitive load between video and audio interpreting conditions. The mean disfluency rate for SI in video condition (*N* = 11) was 19.87 dpm (*SD* = 3.34 dpm, range 14.05 – 25.77 dpm), and the mean disfluency rate for SI in the audio condition (*N* = 11) was 20.18 dpm (*SD* = 4.20 dpm, range 16.29 – 30.97 dpm). Results of the independent sample t-test showed that the disfluency rate in video and audio conditions was not significantly different, $t(20) = -.191$, $p = .850$. The effect size was small ($d = .082$). Although the cognitive load in audio interpreting was slightly higher than that in the video condition, the difference was not statistically significant.

3.2. Gestural behaviors in SI tasks

For gestures, we have identified 317 gestures from SI tasks: 7 deictic gestures, 6 iconic gestures, 57 metaphoric gestures, and 247 beat gestures. Their distribution among fluent and disfluent speech is shown in Table 2:

	Gesture types					
Task types	Deictics	Iconics	Metaphorics	Beats	Total	Percentage
G + Fluency	0	0	0	0	0	0%
G + Quasi-disfluency	5	0	15	90	110	34.7%
G + Disfluency	2	6	42	157	207	65.3%
Total	**7**	**6**	**57**	**247**	**317**	**100%**

Table 2. Number and types of gestures in the SI tasks

It is interesting to notice that all the gestures in SI occurred in disfluent sentences. 65.3% of gestures paralleled disfluencies ('G+ Disfluency'), while the remaining 34.7% was adjacent to disfluencies ('G + Quasi-disfluency'), where disfluencies occurred elsewhere within the sentence and did not overlap with gestures. None of the gestures appeared in a fully fluent sentence ('G + Fluency'). We will further explore the 'G+ Disfluency' category in the next section.

In terms of gesture types, beats were the most frequently used gesture types, which occupied more than three-quarters of the dataset (77.2%), outnumbering all other gesture types. Metaphoric gestures were moderately used with a percentage of 18.1. Deictic (2.2%) and iconic (2.5%) gestures were less frequently used among participants in SI.

Like disfluency rate, gesture rate was calculated as the number of gestures per minute (gpm). The mean gesture rate was 2.25 gpm (*SD* = 2.55 gpm, range 0.16 – 8.87 gpm). T-test showed that the gesture rate in video (*M* = 2.11 gpm, *SD* = 2.49 gpm) and audio (*M* = 2.39 gpm, *SD* = 2.74 gpm) conditions were not significantly different ($t(20) = -.243$, $p = .810$). The effect size was small ($d = .107$). Although participants gestured slightly more during audio interpreting, this difference was not statistically significant, indicating comparable gestural behaviors between video and audio interpreting.

We also tested the correlation between gesture rate and disfluency rate. A Pearson correlation coefficient was computed to measure the linear relationship between gesture rate and disfluency rate. No positive or negative correlations were found between the two variables ($r(20) = .344$, $p = .117$). The effect size was medium ($r^2 = .118$). This indicates that there was no significant linear relationship between gesture and cognitive load in the SI tasks.

3.3. Temporal relation between gesture and disfluency

From the above analysis, we noticed that all the gestures in SI occurred in disfluent sentences, among which 65.3% co-occurred with disfluencies per se. We then focused on this 'G + Disfluency' group, and conducted a detailed analysis of the temporal relation between gesture and disfluency.

There were 207 gestures in this group, including 2 deictic gestures, 6 iconic gestures, 42 metaphoric gestures, and 157 beat gestures. More than half (59.4%, *N* = 123) of these gestures were preceded by disfluencies (the 'Post-disfluency G' category). They overlapped with the first phoneme or word after the disfluency. We have to point out that there was still a partial overlapping between gestures and disfluencies for this category. 79 gestures (38.2%) overlapped with disfluencies per se (the 'Peri-disfluency G' category), and only 5 gestures (2.4%) preceded disfluencies (the 'Pre-disfluency G' category), which means that they overlapped with the first phoneme or word before the disfluency. Like the 'Post-disfluency G' category, these 5 gestures still partially overlapped with the following disfluencies. The distribution of gestures before, within, and after disfluencies is shown in Figure 1.

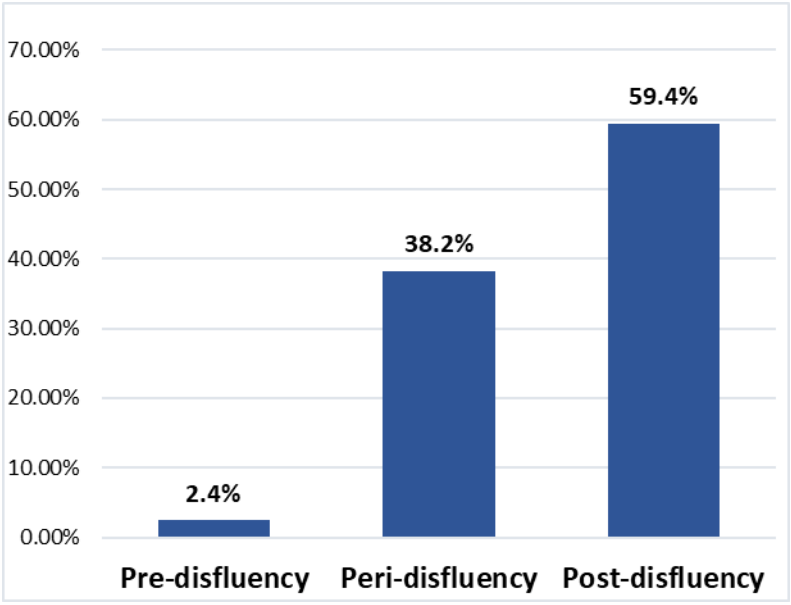


Figure 1. Distribution of gestures before, within, and after disfluencies for the 'G + Disfluency' group (*N* = 207)

We then distinguished whether the disfluency overlapping with gestures was a pause (P) or a false start (FS). We did not differentiate between silent and filled pauses because they had similar functions. 60.9 % of them were pauses and 39.1% were false starts (see Figure 2).

Putting the two figures together, we obtain a more detailed picture, which includes six different sub-categories of the types of gestures and disfluencies, as well as their temporal relations (see Table 3).

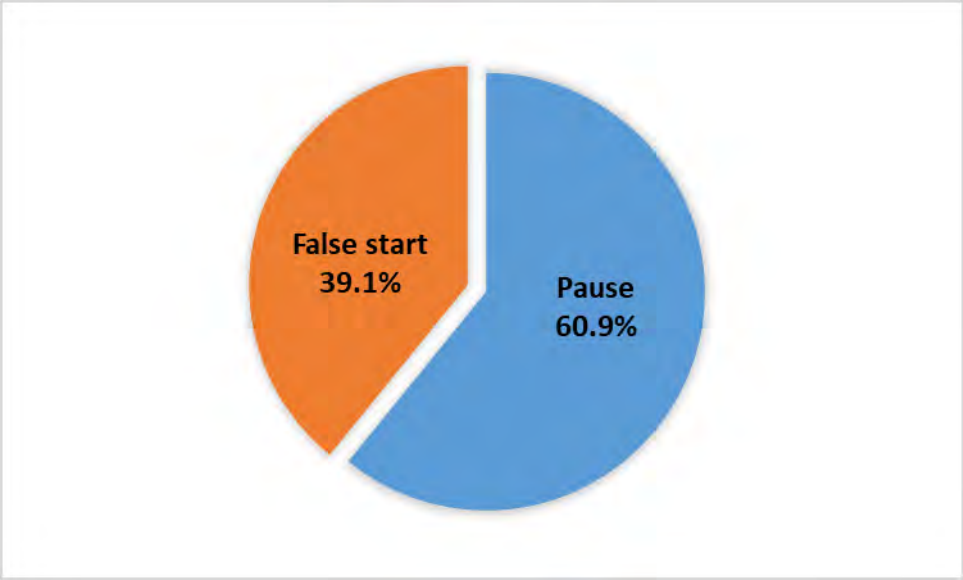


Figure 2. Types of disfluencies overlapping with gestures for the 'G + Disfluency' group (*N* = 207)

	Temporal relations						
	Pre-disfluency G		Peri-disfluency G		Post-disfluency G		
Gesture types	Pre-P	Pre-FS	Peri-P	Peri-FS	Post-P	Post-FS	Total
Deictics	0	0	0	0	0	2	2
Iconics	0	0	0	1	2	3	6
Metaphorics	0	1	10	7	14	10	42
Beats	3	1	42	19	55	37	157
Total	**3**	**2**	**52**	**27**	**71**	**52**	**207**

Table 3. Types of gestures and disfluencies with their temporal relations

Almost all the deictic and iconic gestures occurred *after* disfluency, with only one exception, in which the iconic gesture overlapped with a false start. Only 4 beat gestures and 1 metaphoric gesture occurred *before* disfluency. Metaphoric gestures were evenly distributed between pauses and false starts, where nearly half (*N* = 24) of them were accompanied by pauses, and the other half (*N* = 18) went with false starts. Beat gestures were more closely tied with pauses, with nearly two-thirds (*N* = 100) overlapping with pauses.

4. Discussion

4.1. Relationship between gesture and cognitive load

In addressing the first research question, we will discuss the relationship between gesture and cognitive load in SI from three distinct perspectives.

First, gestures are likely to be produced at moments of processing difficulty. All the gestures in SI were produced in disfluent sentences, among which more gestures were parallel with disfluencies per se (the 'G + Disfluency' group). This result is in line with Stachowiak-Szymczak (2019). In her study, both professional interpreters and trainees produced more gestures when interpreting lists and numbers compared to interpreting narratives. Lists and numbers are problem triggers that require more processing capacity. The result is also partly aligned with that of Adam and Castro (2013) on student interpreters, in which 18.41% of the observed gestures appeared at moments of hesitation in SI, such as pauses or self-correction. This indicates that interpreters are likely to produce gestures when they are experiencing a concrete problem in cognitive processing.

Second, interpreters tend to gesture more in tasks with a higher cognitive load. When comparing different task conditions, disfluency rate in audio interpreting was found to be slightly higher than that in video interpreting. The same trend was found in the comparison of gesture rate between the two conditions. Even though the differences in both cases were not statistically significant, more gesture was used in the condition with a higher cognitive load. The difference of disfluency rate could be explained by the effect of codes and modalities on task interference (Seeber, 2007; 2011). In the audio condition, the input and output processes are both in auditory modality and only verbal processing is involved; while in the video condition, the level of processing underlying visual modality is multimodal and is different from that of auditory modality. Thus, greater interference and more cognitive load would arise in audio interpreting than in video interpreting, because the two processes in audio interpreting have common structures. Furthermore, the fact that more gestures are used in audio interpreting corroborates the facilitative role of gestures (Kita *et al*., 2017). Gesturing while speaking can reduce the cognitive load on working memory (Goldin-Meadow *et al*., 2001).

Third, gestures were embodied, multimodal manifestations of exported load. In our dataset, we found that the use of gestures was parallel with or adjacent to processing difficulty: all the gestures are produced in disfluent sentences. However, the correlation between gesture and cognitive load did not reach a significant level. This could be explained by the notion of 'exported load' (Gile, 2008), which refers to the phenomenon that extra cognitive load may be carried over to downstream segments, leading to cognitive saturation at a later stage. Based on the assumption that speech and gesture are generated from the same cognitive mechanism (Hostetter & Alibali, 2008), exported loads may be reflected in speech as disfluencies or as speech-accompanying gestures.

For the 'G + Disfluency' group, where gestures were in parallel with disfluencies, more than half (59.4%) were produced after disfluency (the 'Post-disfluency G' category). Here, disfluency reflects a moment at which a processing problem arises and the interpreter stops the speech production process to solve the problem. Solving problems requires more effort, leading to an increase in local cognitive load. When the problem is solved and speech production is resumed, the extra load is carried over downstream, which might be manifested in the gestural modality. Thus, for gestures parallel with disfluencies, they tend to appear after disfluency.

For the 'G + Quasi-disfluency' group, gestures could also reflect exported load. These gestures are performed with disfluent sentences, but they are not in parallel with disfluencies per se. Loads from the previous sentence(s) could influence the downstream processing, resulting in disfluencies or gestures.

4.2. Functions of gesture under cognitive load

For the second question, we will discuss the function of each gesture type under cognitive load.

First, in our dataset, the types of gestures produced during silence, i.e. produced with pause, are beats and metaphorics. In the 'Peri-P' sub-category, which refers to gestures parallel with pauses, there are 10 metaphoric gestures and 42 beat gestures. No deictic and iconic gestures are found in this sub-category. Through detailed analysis of these gestures, we find that their usage is in line with the cognitive functions described by other researchers. In the following examples, we followed McNeill's (1992) transcription system to transcribe gestures. For disfluencies, silent pauses are transcribed as a double slash (//) plus their duration in parentheses. When presenting examples, we include both SL and TL as well as a word-for-word English equivalence in italicized form beneath the corresponding character in TL.

The use of metaphoric gesture during silence echoes McNeill's (1985) description of 'silent gesture': such gesture relies on the conceptualization of linguistic units as containers. Example (1) shows the use of a silent metaphoric gesture by participant No. 7 (P07) when interpreting speech A.

Example (1): Silent metaphoric gesture in audio interpreting

SL (A): What you find are three clusters of character strengths.

TL (P07): // (0.7s) 你会发现 **[… // (0.8s)]** **Metaphoric** [三] Beat [种] Beat [不同的] Beat 评价

you will discover *three* *kinds* *different* *evaluations*

In this example, after interpreting the first three words, the participant stopped for 0.8 seconds, and a metaphoric gesture was produced during this silent pause. The form of this gesture is a small finger-lift movement (Cienki, 2021; this volume), where both of her thumbs were raised upward and her right forefinger was stretched forward (see Figure 3I). This is a typical gestural form to present a point when the speaker is seated, with hands on a table in the front, palms facing the speaker (Cienki, 2021, p. 20). Then the participant resumed interpreting and produced three beat gestures, which were miniature tapping movements performed by her right forefinger, along with the subsequent verbal products.

Figure 3. The silent metaphoric gesture in audio interpreting by P07

A plausible explanation was that the participant used a waiting strategy (Seeber, 2011), by which she halted TL production to wait for more input from SL. At the same time, the metaphoric gesture called forth a container in which the meaning of the subsequent segment could be filled. This gesture was not semantically related to the verbal product, but it reflected the problem-solving strategy during silence (Kita *et al*., 2017).

The use of beat gestures during silence is also in line with McNeill's (1985) description: they serve as an attempt to get the speech process going again. Lucero *et a*l. (2014) also found that

beat gestures can facilitate speech production. Example (2) illustrates the use of silent beat gestures by participant No. 2 (P02) when interpreting speech A.

Example (2): Silent beat gestures in video interpreting

> SL (A): So these can be called, in David Brooks' language, the "résumé strengths", because these are the things that get you hired.
>
> TL (P02): // (2.6s) <呃> 我的朋友 / David Brooks **[[…] […] // (7s)]** $_{\text{Beat}}$ **[**嗯**]** $_{\text{Beat}}$ / 他
> *er my friend David Brooks en he*
>
> 认为这是一个长期形成的过程 // (2.6s)
> *thought this is a long formation process*

There was a long silent pause in this segment, which lasted for 7 seconds. During the pause, the participant produced two beat gestures, which were downward movements performed by her right thumb (see Figure 4). Each gesture stroke is marked by a pair of square brackets in transcription.

Figure 4. The silent beat gesture in video interpreting by P02

In the retrospection comment, the participant mentioned that she could not fully understand the meaning of the phrase "résumé strengths" and was searching for a Chinese equivalence, even though she did understand the meaning of the individual word "résumé". She did not come up with an equivalent expression in the TL, and then she resumed interpreting the subsequent segment with a filler word "嗯", which was marked by a third beat gesture.

The two gestures produced during the silent pause also reflected, in an indirect way, the problem-solving strategy. The number of gestures indicates that the participant tried to resume her interpretation twice when the production of interpreting was halted. The participant's retrospection echoes the finding of Adam and Castro (2013). When beat gestures were produced at moments of hesitation, the interpreter was either having a comprehension problem or engaging in word search.

The third gesture produced with the filler word indexed the end of silence and the beginning of the following interpretation. These three gestures are not semantically related to verbal products, but they could serve an emphasizing function in that they implicitly contrast the absence of a word with its desired presence (McNeill, 1985, p. 359). In this sense, silent beat gestures could have a meta-cognitive function, namely monitoring.

Second, deictic and iconic gestures are not produced before the semantically related linguistic items but rather parallel with them. In McNeill's (1985) illustration of gestural anticipation, a gesture could be produced prior to the production of its corresponding linguistic item. However, in our dataset, we found that deictic and iconic gestures were produced with the corresponding word itself. The anticipation of gestures is not obvious. Example (3) shows the use of iconic gestures by participant No. 1 (P01) when interpreting speech B.

Example (3): Iconic gesture in video interpreting

SL (B): When the work came back, I calculated grades.

TL (P01): // (0.6s) 之后 / **[**收回来**]** $_{\text{Iconic}}$ 这些任务之后呢我开始去算他们的分数 // (1.2s)

then collect back these tasks after I began to calculate their grades

Figure 5. The iconic gesture in video interpreting by P01

When producing the phrase "收回来", the participant produced an iconic gesture, which was a backward movement toward her own body, mimicking the action of "collect" (see figure 5). This gesture was semantically redundant, for it expressed the same linguistic meaning with speech. In terms of temporal relations, the gesture is produced simultaneously with its linguistic counterpart, not preceding it.

Deictic and iconic gestures are semantically related to concurrent speech (Arbona *et al*., 2023). There were only 2.2% (*N* = 7) deictic gestures and 1.9% (*N* = 6) iconic gestures in the dataset, including the one in Example (3), and all were produced with linguistic items. However, in similar works, simultaneous interpreters produced more of such semantically related gestures (Martín de León & Fernández Santana, 2021; Zagar Galvão, 2020). This could be influenced by factors such as the content of the speech, the speaker's style, and cultural differences.

The lack of anticipation for semantically related gestures could be explained by the task peculiarity of SI. Unlike in spontaneous speech production, the meaning of the interpreting product comes not from the interpreter, but from the speaker. The interpreter has to receive, and perhaps wait for, the input from the speaker, while producing the output in another language. This mental process is different from that of gestural anticipation in spontaneous speech. Given that the cognitive demands in SI are high, there is not enough working memory capacity to store the pre-activated representations for gestures. Such representations are only activated when producing the verbal product and simultaneously produced in gestures. The production of deictic and iconic gestures could be affected by cognitive load.

5. Conclusion

This study explored the relationship between gesture and cognitive load in simultaneous interpreting. Our findings are as follows. First, gestures in SI are likely to be produced with processing difficulty, especially when interpreters are experiencing a concrete problem. Even if the correlation between gesture and cognitive load is not statistically significant, interpreters tend to gesture more in tasks with a higher cognitive load. This corroborates the facilitative role of gestures in cognitive processing. Based on the assumption that speech and gesture are generated from the same cognitive mechanism, gestures could be an embodied, multimodal manifestation of 'exported load'. The temporal relations between gesture and disfluency give support to this claim. Processing difficulty could result in a speech breakdown and a parallel gesture is produced. The strokes of this gesture tend to follow the speech disfluency.

Additionally, processing problems could also influence downstream segments, leading to both disfluencies and gestures. However, in such cases, gesture and disfluency co-occur within a sentence boundary, but they do not overlap. Both of the two modalities could be reflections of exported loads. Second, beat and metaphoric gestures are connected with processing difficulty. Silent beat gestures could reflect the interpreter's monitoring of the interpreting process, which means that beats might have a meta-cognitive function. On the other hand, silent metaphoric gestures may reflect the problem-solving strategy. When producing metaphoric gestures, interpreters stopped TL production and called forth a container for the subsequent speech segments to fill in. When speech production comes to a halt, cognitive processing does not stop. Silent gestures are embodied manifestations of such processes. Third, deictic and iconic gestures could convey concurrent semantic meaning, but such gestures are less frequently used by the interpreters of the study. Due to the task peculiarity of SI, there is a lack of anticipation for such gestures. Their production could be affected by cognitive load. Semantically related gestures provide a multimodal perspective for studying the cognitive processes of interpreting.

This study contributes to the understanding of SI as an embodied, multimodal cognitive activity. The findings have indicated several lines of research in the future. There is a difference between the number of gestures produced by interpreters in this study and by interpreters from other countries. A comparative study could unveil the cultural differences behind gestural styles. Future research could also make comparisons between gestures in SI and CI to further explore the relationship between gesture and cognitive load. One limitation of this study is the number and the level of professional competence of participants. In the future, we will recruit more participants as well as professional interpreters. A comparison between professional and student interpreters will also shed light on the effect of interpreting competence on gestural behaviors.

6. Acknowledgement

This research is supported by "the Fundamental Research Funds for the Central Universities" in UIBE (23QD12). We would like to thank professor Alan Cienki, Dr. Celia Martín de León, Dr. Sílvia Gabarró-López and the two anonymous reviewers for their valuable comments and suggestions on earlier drafts of this paper. The responsibility for the content and any remaining errors remains exclusively with the authors.

7. References

Adam, C., & Castro, G. (2013). Schlaggesten beim Simultandolmetschen – Auftreten und Funktionen, *Lebende Sprachen, 58*(1), 71–82. https://doi.org/10.1515/les-2013-0004

Arbona, E., Seeber, K., & Gullberg, M. (2023). Semantically related gestures facilitate language comprehension during simultaneous interpreting. *Bilingualism: Language and Cognition*, *26*(2), 425–439. https://doi.org/10.1017/S136672892200058X

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bóna, J., & Bakti, M. (2020). The effect of cognitive load on temporal and disfluency patterns of speech: Evidence from consecutive interpreting and sight translation. *Target*, *32*(3), 482–506. https://doi.org/10.1075/target.19041.bon

Chen, S. (2017). The construct of cognitive load in interpreting and its measurement. *Perspectives*, *25*(4), 640–657. https://doi.org/10.1080/0907676X.2016.1278026

Christoffels, I., & De Groot, A. (2005). Simultaneous interpreting: A cognitive perspective. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 454–479). Oxford University Press.

Cienki, A. (2021). From the finger lift to the palm-up open hand when presenting a point: A methodological exploration of forms and functions. *Languages and Modalities*, *1*, 17–30. https://doi.org/10.3897/lamo.1.68914

Cienki, A., & Iriskhanova, O. K. (2020). Patterns of multimodal behavior under cognitive load: An analysis of simultaneous interpretation from L2 to L1. *Voprosy Kognitivnoy Lingvistiki*, *1*, 5–11. https://doi.org/10.20916/1812-3228-2020-1-5-11

Gile, D. (1999). Testing the Effort Models' tightrope hypothesis in simultaneous interpreting – A contribution. *Hermes*, *23*, 153–172. https://doi.org/10.7146/hjlcb.v12i23.25553

Gile, D. (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *Forum*, *6*(2), 59–77. https://doi.org/10.1075/forum.6.2.04gil

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychological Science*, *12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Gumul, E. (2021). Explicitation and cognitive load in simultaneous interpreting: Product- and process-oriented analysis of trainee interpreters' outputs. *Interpreting*, *23*(1), 45–75. https://doi.org/10.1075/intp.00051.gum

Han, C., & An, K. (2020). Using unfilled pauses to measure (dis)fluency in English-Chinese consecutive interpreting: In search of an optimal pause threshold(s), *Perspectives*, *29*(2), 1–17. https://doi.org/10.1080/0907676X.2020.1852293

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059

Litvinenko, A. O., Kibrik, A. A., Fedorova, O. V., & Nikolaeva, J. V. (2018). Annotating hand movements in multichannel discourse: Gestures, adaptors and manual postures. *The Russian Journal of Cognitive Science*, *5*(2), 4–17.

Lucero, C., Zaharchuk, H., & Casasanto, D. (2014). Beat gestures facilitate speech production. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 898–903). Curran Associates.

Martín de León, C., & Fernández Santana, A. (2021). Embodied cognition in the booth: Referential and pragmatic gestures in simultaneous interpreting. *Cognitive Linguistic Studies*, *8*(2), 277–306. https://doi.org/10.1075/cogls.00079.mar

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*(3), 350–371. https://doi.org/10.1037/0033-295X.92.3.350

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.

Muñoz Martín, R. (Ed.). (2016). *Reembedding translation process research*. John Benjamins.

Seeber, K. G. (2007). Thinking outside the cube: Modeling language processing tasks in a multiple resource paradigm. *Interspeech 2007, 8th Annual Conference of the International Speech Communication Association* (pp. 1382–1385). https://doi.org/10.21437/Interspeech.2007-21

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories - new models. *Interpreting*, *13*(2), 176–204. https://doi.org/10.1075/intp.13.2.02see

Seeber, K. G. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target*, *25*(1), 18–32. https://doi.org/10.1075/target.25.1.03see

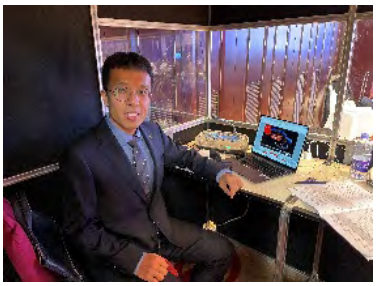
Seeber, K. & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, *16* (2), 228–242. https://doi.org/10.1177/1367006911402982

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. *Proceedings of the 6th international conference on language resources and evaluation (LREC 2008)* (pp. 816–820).

Stachowiak-Szymczak, K. (2019). *Eye movements and gestures in simultaneous and consecutive interpreting*. Springer.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. https://doi.org/10.3758/BF03196322

Zagar Galvão, E. (2020). Gesture functions and gestural style in simultaneous interpreting. In H. Salaets & G. Brône (Eds.), *Linking up with video: Perspectives on interpreting practice and research* (pp. 151–179). John Benjamins. https://doi.org/10.1075/btl.149.07gal



Yuetao Ren
University of International Business and Economics
Huixin Dongjie, 10
100029 Beijing
China

renyuetao@gmail.com

Biography: Yuetao Ren is a lecturer in Translation and Interpreting at the University of International Business and Economics (UIBE), China. He received his PhD degree in Interpreting Studies from Renmin University of China. His research interests include interpreting studies, gesture studies, and multimodality. He also works as the coordinator for the conference interpreting program at UIBE and as a freelance conference interpreter.

Jianhua Wang
Renmin University of China
Zhongguancun Street, 59
100872 Beijing
China
wjhsfl@ruc.edu.cn

Biography: Jianhua Wang is a professor of Translation and Interpreting at Renmin University of China (RUC). He holds a PhD degree in Cognitive Psychology. His current research interests include the cognitive process of translation and interpreting, translation and communication, multimodal translation, discourse studies, as well as global and area studies. He also serves as the associate dean of the School of Foreign Languages at RUC.