ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 14, No. 6, pp. 1757-1770, June 2024
DOI: https://doi.org/10.17507/tpls.1406.16

Delving Into the Influence of Visual Input on the Load-Related Silent Pauses During SI: What About Gestures?

Yunxiao Jiang
Faculty of Languages and Translation, Macao Polytechnic University, Macau, China

Lili Han[*]
Faculty of Applied Sciences, Macao Polytechnic University, Macau, China

Yuqi Sun
Faculty of Arts and Humanities, University of Macau, Macau, China

***Abstract*—This research delves into the realm of simultaneous interpreting (SI) with a focus on the Portuguese-Chinese language pair, examining the interplay between visual inputs and cognitive load. This study posits that visual cues such as hand gestures may influence the cognitive load during SI, a topic that remains controversial in interpreting studies. To address this, we conducted an empirical study involving 18 trainee interpreters divided into two groups: a control group receiving only audio input and an experimental group with additional video input. Utilizing ELAN 6.3 software, we analyzed silent pauses exceeding 300ms to gauge the cognitive load. The research focused on how audio and video inputs impact these silent pauses, with a special emphasis on segments accompanied by semantically related hand gestures. The results revealed that the average duration of silent pauses was marginally shorter for interpreters with video input, although the differences between the two groups were not statistically significant. Intriguingly, for both groups, the duration of pauses significantly increased during segments with semantically related gestures, underscoring the inherent high cognitive demand of these segments, irrespective of visual input. A notable discovery was the marked increase in fluency for participants with visual access when interpreting segments accompanied by gestures, which suggests that semantically related gestures provide cognitive benefits. Overall, this study contributes to the ongoing discourse on the role of visual inputs in SI, highlighting the potential of gesture input to alleviate cognitive load and improve interpreter performance.**

***Index Terms*—visual input, semantically related gestures, silent pauses, SI, cognition**

I. Introduction

As a cognitive activity of meaning-making for communication, SI has been increasingly understood as a multimodal activity (Mikkelson & Jourdenais, 2015, p. 293). There is evidence that the integration of information from multiple senses helps brain to disambiguate (Koelewijn et al., 2010; Sumby & Pollack, 1954), speed up the response to the target stimuli (Miller, 1982; Molholm et al., 2006; Teder-Sälejärvi et al., 2002), promote temporary storage of information (Baddeley, 2000), and result in a better recall afterwards (Gieshoff, 2018, p. 31). According to Gieshoff (2018), these facilitating effects of “redundant” (Seeber, 2017, p. 464) multisensory information only exist when the stimulus from both sensory channels (audio and video) are spatially, temporarily and semantically congruent. However, the multiple cues received by the interpreters during the SI task may also compete for same pool of cognitive resources (Seeber, 2007), and cause an increasement of cognitive effort especially during the stage of perception and cognition (Prandi, 2023; Seeber, 2017). By now, there are two opposite hypotheses under debate: 1) since multimodal input benefits language comprehension in L1 and L2, will it (in both acoustic and visual channel) reduce the cognitive load of interpreter due to the multisensory binding effect (Gieshoff, 2018, p. 30)? Or 2), on the contrary, in light of the Cognitive Resource Footprint (CFR) of Seeber (Seeber, 2011, 2012), will it cause increasing cognitive load, or even overload?

Although some empirical studies were conducted to reveal the impact of visual input on simultaneous interpreting, their results were inconsistent due to differences in object of study (lip movements, visual presentations, gestures, etc.) and measurement used to estimate cognitive load (cognate translations, silent pauses, eye movements, etc.) involved in task, which calls for further empirical studies to test the impact of different types of visual cues on the cognitive load and performance of interpreters. Rennert (2008), for example, reviewed types of visual input that can possibly influence the SI task and carried out an experiment. The results suggested no appreciable positive or negative effect of visual

[*] Corresponding Author.

input, though some interpreters benefit from the additional or redundant information provided by the visual cues. In addition, the perception of interpreters in SI tasks was also studied to evaluate the influence of visual access. Moser-Mercer (2005), for example, conducted a study to investigate the parameters that influence the feelings of presence among professional interpreters in remote interpreting settings. The results of her meta-analysis indicated that the view of the speaker was the only positive value of remote interpreting that was announced by the interpreters, while other factors such as the view of the conference room, alienation from the conference room, and motivation did not have a significant impact on their feelings of presence (p. 733). In a study by Baxter (2016), the effect of using visual aids in interpreting was examined. Contrary to the initial hypothesis, the ear-voice span significantly increased with the use of visual presentation. Gieshoff (2017) examined the impact of lip movements on the cognate translation, which refers to words that share the same etymological roots in two languages and whose orthographic and phonetic representations significantly overlap (p. 316), for example, "configuration" in English and "configura ção" in Portuguese, as an indicator of cognitive load experienced by the interpreter. The study explored how the presence or absence of white noise in the source speech affected the interpretation, revealing an increase in cognate translations when no lip movements were present, particularly when white noise was introduced, implying a higher cognitive load without the presence of visible lip movements of the speaker. The same author (Gieshoff, 2021a) also tested the duration of silent pauses when the interpreters are (not) exposed to lip movements of the source speech speaker. The results indicate the interpreter benefits from the visual input, with significantly shorter pause duration compared with the control group, implying also the positive cognitive influence of lip movements. Previous studies on different language pairs such as German and English (Rennert, 2008) and Greek and English (Gieshoff, 2018), and to the best knowledge of the author, little research has been conducted concerning the Portuguese-Chinese language pair about the synergy between visual inputs and cognitive load, except the most recent one about the moment analysis by Han et al. (2023) in light of translanguaging theory and of the complex, dynamic system theory (CDST) approach, which focused on the process of interpreting, examining the workflow tasks of CI and SI. However, our present study is both process- and product-oriented, aiming to bridge the gap in academia, by further exploring the influence of visual input on the load-related disfluent performance during SI by delving into the silent pauses.

II. Silent Pauses as Indicator of Cognitive Fluency in SI

In interpreting studies, disfluencies reflect the difficulty of the source text, such as syntactic complexity (Shen et al., 2023), dependency distance (Jiang, 2020), informational load (Kajzer-Wietrzny, 2023), or lexical density. In fact, disfluencies such as hesitations (including lengthening of vowels and filled pauses), silent pauses with extend of more than 0.3 second and interruptions of expressions (including repetitions, false starts and self-correction) have been considered as indicators of cognitive load during interpreting tasks in previous studies (Mead, 2000; Skehan, 2003; Song, 2020), as they occur when the interpreter processes complex or unfamiliar information and needs more time or attention to produce a coherent output (Jiang & Jiang, 2020; Plevoets & Defrancq, 2016). In addition, disfluencies (especially silent pauses) may indicate the cognitive strategies that interpreters use to cope with high load, such as simplification, segmentation, anticipation, or monitoring (Zhao, 2022). To study the different cognitive load imposed on the interpreter by different types of dependency distance (long or short max dependency distance) of source speech, Jiang and Jiang (2020) examined the four categories of disfluencies (cf. Shreve et al., 2011), namely, silent pauses, filled pauses, repetitions and self-corrections. The researchers observed a significant higher frequency of disfluencies under the condition of long max dependency distance.

Among the various types of disfluent phenomenon, silent pauses are widely examined in previous studies regarding the cognitive load in simultaneous interpreting. Apart from Jiang and Jiang (2020) mentioned earlier, Gieshoff (2021b) also examined the duration of load-related silent pauses depending on the interpreter seeing or not seeing the lip movement of the source-text speaker. The researcher revealed that the duration of the silent pauses in interpreting was significantly shorter when the interpreter sees the lip movement of the source text speaker. Song and Li (2020) also examined various types of disfluencies including silent pauses in SI, indicating that the mean duration of silent pauses was shorter in the SI output of trainee interpreters with higher lexical retrieval efficiency. This finding implies that trainee interpreters with better cognitive capacity achieve better cognitive fluency during the SI tasks, which is observable through shorter silent pauses in their interpreting. Moreover, B. Wang and Li (2015) probed into the characteristics of and motivations for pauses, which offers the closest relevance to our study. Their study, which specifically examines the language pair of English and Chinese in simultaneous interpreting, offers valuable insights into the characteristics of pauses. The findings reveal that pauses in the target output speech occur less frequently compared to the source input speech but have longer duration. Furthermore, the pauses in the target output speech occur frequently at the moment of interpreting the source input speech with syntactic complexity, which indicates a correlation between language structure and pause during the process of interpretation.

Though some studies consider only unusually long pauses that lead to audience’s discomfort are disfluencies in SI (Song, 2020), shorter disfluencies, ranging from 200ms to 300ms, are commonly employed to study the synergy between delivery of speech and cognitive fluency, and perceived as measure of cognitive fluency (Kahng, 2014; Xin, 2020). In the range of studies we have examined regarding various focus including the spontaneous speech in first and second languages, consecutive and simultaneous interpreting, the majority have identified a threshold for silent pauses that falls between 0.25s to 0.4s. A seminal study by Goldman-Eisler (1958) proposed a threshold of 0.25 seconds for silent pauses and identified that a majority (71.5%) of pauses in oral speech fell within this time range. This standard was subsequently adopted in research conducted by Grosjean and Deschamps (1972). Raupach (2011) defined pause at intervals of 0.3 seconds or longer, either within or between sentences. Delineating hesitation from pauses, Riggenbach (1991) put forth distinct thresholds: 0.2 seconds for micro-pauses, a 0.3 to 0.4 second range for hesitation, and an unfilled pause spanning between 0.5 to 3 seconds (Goldman-Eisler, 1958). Other studies, such as Towell et al. (1996) and Mead (2000), suggested a minimum cut-off point around 0.28 seconds and 0.25 seconds respectively, with Mead setting an upper cut-off at 3 seconds (X. Wang & Wang, 2022).

Table 1 exhibits a variety of antecedent studies pertaining to different language-related domains and their respective thresholds for silent pauses. In consideration of the preceding researches on cognitive and utterance fluency, the present study adapts a threshold put forth by Yang et al. (2020). Their work, which explores the cognitive load in SI with text, aligns closely with the study object of our current investigation. Thus, in the context of simultaneous interpreting, we have established the minimum silent pause threshold for SI production to be demarcated at 0.3 seconds.

TABLE 1
THRESHOLD ADOPTED IN PREVIOUS STUDIES

Font	Area of study	Types of discourse	Silent pauses threshold
(Kahng, 2014)	Second language acquisition	Oral production tests for English Foreign Language speakers	0.25s
(Xin, 2020)	Interpreting studies	Consecutive interpreting between Chinese and English	0.3s
(Hieke, Kowal, & O'Connell, 1983)	First language discourse	Political discourse in English and French	0.13s
(Yang, 2019; Yang et al., 2020)	Interpreting studies	SI with texts	0.3s

To sum up, previous empirical studies examined the cognitive load of SI tasks under manipulated conditions and manifested that different levels of cognitive load can be imposed on interpreters. The differentiated cognitive loads are reflected on the duration of silent pauses in SI output, and the pause duration, in its turn, was employed to examine cognitive loads. Moreover, the method by applying the pause duration as measurement permits a real work scenario of the interpreters, unlike the intervention of measuring devices such as eye-tracking tools or EEG headset. Therefore, we decided to adopt silent pause duration as indicator of the cognitive fluency of the trainee interpreters in the current study.

III. EXPERIMENT

We enlisted 18 participants – trainee interpreters, all with one year of interpreting training experience, from the master’s program in Chinese and Portuguese Translation and Interpreting from a university of Macau. These trainee interpreters are native speakers of Chinese with C1 proficiency level in Portuguese as second language and all signed informed consent forms to participate in the study.

The 18 participants were equally divided into two groups, control group and test group. The test group, composed of nine trainee interpreters, was tasked with performing simultaneous interpreting (SI) both from Chinese to Portuguese and from Portuguese to Chinese, with video input. In contrast, the control group, composed of the remaining nine trainee interpreters, carried out the same tasks but without access to video input.

To facilitate potential preparatory work by our trainee interpreters, we announced the discourse topic three hours prior to the commencement of the experiment. Moreover, for precise timeline alignment of the semantic gestures in the source speech, the interpreting performance' pauses, and the interpretation of the gestured source speech by the interpreters, we marked the start time of the source speech in the audio recording of the interpreting.

The source speech videos in Chinese and Portuguese were deliberately selected and edited to guarantee a similar frequency of the occurrence of gestures semantically related to the source speech. Both speeches are related with the socio-cultural aspects of reading habits: the Chinese speech has 36 semantic gestures and 48 beat gestures, while the Portuguese speech has 34 semantic gestures and 36 beat gestures in the edited video clip. The source speakers ́hand gestures semantically related to the verbal language were marked on ELAN 6.3 (Wittenburg et al., 2006) for analysis, and the interpreting performance's silent pauses over 0.3s, detected and extracted first automatically and further revised by human researcher, were also imported into ELAN 6.3 (Wittenburg et al., 2006). Hence for each SI file, we have the silent pauses with start time and end time marked and the beginning and end of the gesture-accompanied source-speech segments. Apart from those, we also have the silent pauses of the critical segments for each trainee interpreters marked in the eaf. files of ELAN (Figure 1).

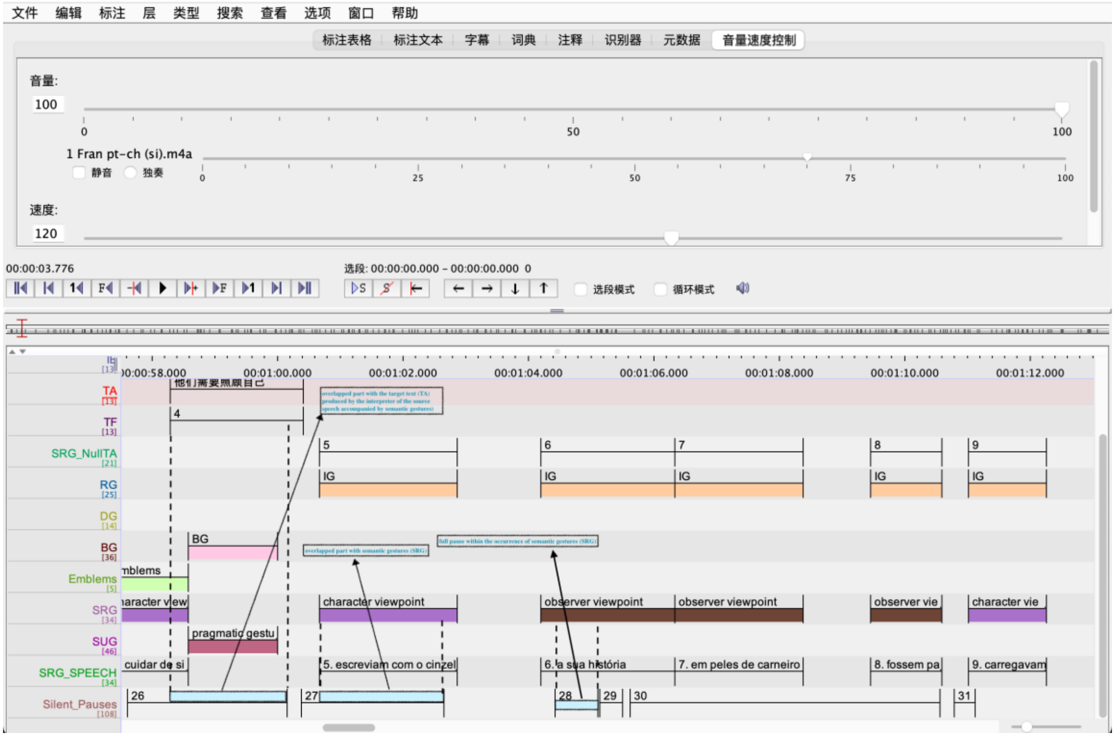

Figure 1. Screenshot for ELAN Layout

As shown in Figure 2, two kinds of the silent pause duration were calculated for the present analysis:

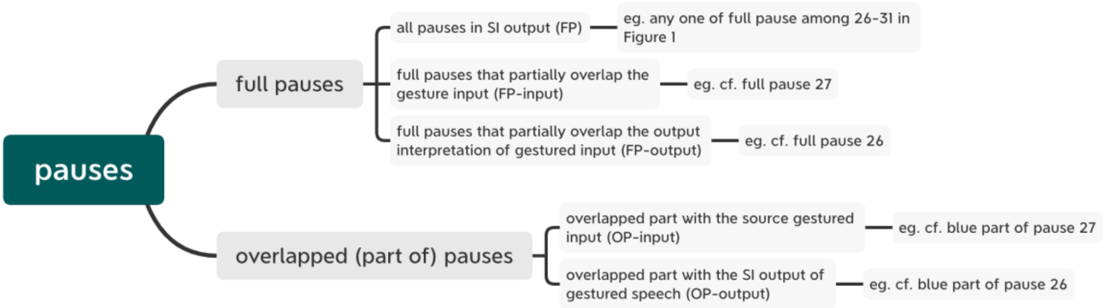
pauses
full pauses
all pauses in SI output (FP)
eg. any one of full pause among 26-31 in Figure 1
full pauses that partially overlap the gesture input (FP-input)
eg. cf. full pause 27
full pauses that partially overlap the output interpretation of gestured input (FP-output)
eg. cf. full pause 26
overlapped (part of) pauses
overlapped part with the source gestured input (OP-input)
eg. cf. blue part of pause 27
overlapped part with the SI output of gestured speech (OP-output)
eg. cf. blue part of pause 26

Figure 2. Types of Pauses for Analysis

By choosing FP and OP we aim to collect comprehensive data, to facilitate the latter data analysis from a comparative perspective.

For the data analysis, we used non-parametric analysis for the comparison of pause durations for different groups. Mann Whitney U test (McKnight & Najab, 2010) was used for the two sided test of mean difference comparison and for the monotonic analysis (see Table 2). We also adopted the Random Forest (RF) Regressor (Grömping, 2009), as a non-parametric regression analysis model, which permits a simultaneously processing of both categorical features (such as the 'input condition', 'direction', and 'different gesturing condition of the source speech' considered in this study) and quantitative features (like the 'pause duration' in this study) (see Table 3). The random forest (RF) model facilitates the computation of feature importance, helping to understand which characteristics lead to a significant impact by predicting the dependent variable's value though the regression model. In the present study, instead of relying on the regression model to predict the values of 'pause duration', we seek to understand, through the process of model fitting, which factors (such as 'direction', 'different gesturing condition of the source speech') have a more substantial impact on 'silent pause duration'. Beyond calculating feature importance, we also utilized partial dependence plots (PDP), a diagram about statistical data, to visualize the relationship between specific features and prediction outcomes. Partial dependence plots can demonstrate the influence of a particular variable on the dependent variable, assuming all other independent variables are held constant.

As shown in Table 2, different types of FP or OP were compared and analyzed for the SI output audios of the control group, the experimental group, the Chinese to Portuguese direction, the Portuguese to Chinese direction.

TABLE 2
TYPES OF PAUSES EXAMINED FOR DIFFERENT PURPOSE WITH DIFFERENT DATASET USING MANN WHITNEY U TEST

Compared condition	Examined silent pause types	Examined groups
Audio input vs. video input	FP	Two directions separately
PT to CH vs. CH to PT	FP	overall
FP-input vs overall condition	FP-input vs. FP	All input conditions and directions separately
FP-output vs overall condition	FP-output vs. FP	All input conditions and directions separately
Audio input vs. video input	OP-input	Two directions separately
Audio input vs. video input	OP-output	Two directions separately

TABLE 3
TYPES OF PAUSES EXAMINED FOR DIFFERENT PURPOSE WITH DIFFERENT DATASET USING RANDOM FOREST REGRESSION ANALYSIS

Purpose	Group	Pause types
Impact of each feature	All audios	FP+OP
Impact of directionality controlling input condition	Only audio input	FP+OP
	Only video input	FP+OP
Impact of input condition controlling SI directionality	Only CH to PT	FP+OP
	Only PT to CH	FP+OP

IV. DATA ANALYSIS AND RESULTS

We first carried out a pre-processing of data, by examining the data's distribution. The findings showed that the pause duration is still not normally distributed even after the log-transformation, box-cox transformation, and square transformation, so we used a non-parametric test, that is Mann Whitney U test to compare the differences between mean pauses duration (MPD) for different groups of interpretation. Figure 3 demonstrates the distribution of the duration of all pauses over 300ms extracted from our interpreting performance audios, from which we can know directly that the data is not normally distributed.

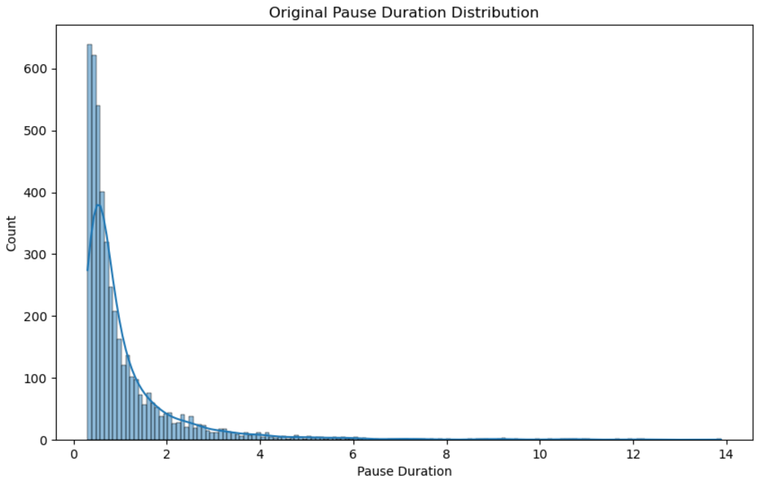

Figure 3. Distribution of Silent Pauses Duration

A. *Comparison Between the Two Input Conditions of SI*

We then analyzed the difference of mean silent pause duration between the video and the audio input condition. It is found that the mean silent pause is slightly shorter for the interpreting of the video input group. However, as shown in Table 4, the difference between the two is not statistically significant, with p-value in both directions of interpretations above 0.05.

TABLE 4
DIFFERENCE ON PAUSE DURATION BETWEEN VIDEO AND AUDIO INPUT GROUP

Direction	FP duration for audio input group	FP duration for video input group	p-value	Test used
CH-PT	0.95	0.92	0.906	Mann Whitney U
PT-CH	1.39	1.34	0.385	Mann Whitney U

As shown in the boxplot, another diagram of statistical data, the mean duration of silent pauses does not significantly distinguish the interpreting performance of the video input group from that of the audio input group.

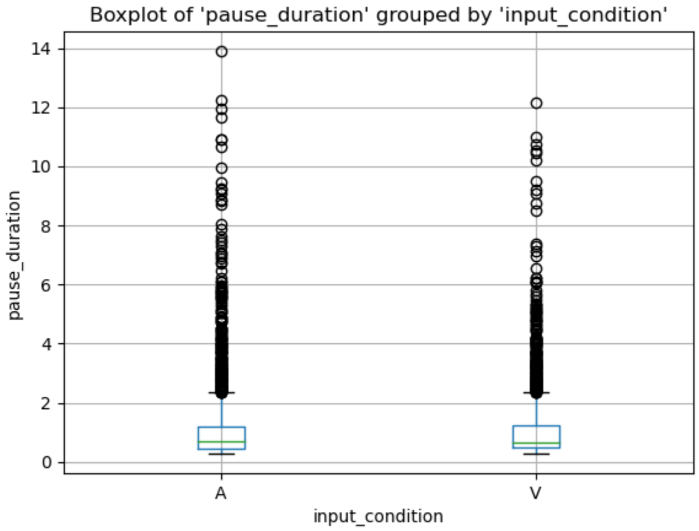

Figure 4. Comparison of Silent Pauses Duration for Video and Audio Input Groups

To discern the differentiated impact of the audio input and of visual input that might shed light on the cognitive processing involved, we further investigated one specific kind of visual input – the semantically related gestures. We examined the FP-input, that is, the silent pauses that concur at the speech gestured segments (pause 27 in Figure 1) and compared these to the overall duration level of pauses in the whole SI audio. The findings indicated that the mean duration of these pauses exceeded the overall average of pauses, not only across both directions of interpretation, but also under both video and audio input conditions.

As shown in Table 5, the differences are statistically significant and the results revealed that when the interpreter hears (or sees) the speech segments accompanied by the semantic gestures of the source-speech speaker, a significant longer pause occurs in their oral production, indicating the over cognitive load of the participants while trying to perceive this part of source speech. The findings suggest that during the perceptual/cognitive stage of SI activity, the presence of semantic gestures made by the speaker alongside the source speech imposes additional cognitive load on trainee interpreters. Consequently, this increased cognitive load incurred from both auditory and visual stimuli leads to a deterioration in the quality of online SI output for both experimental and control groups.

TABLE 5
FP-OUTPUT DURATION VS. OVERALL MEAN PAUSE DURATION

Direction	Input Condition	Duration for all FP	Duration for FP-input	p-value	Test used
CH-PT	audio	0.95	1.19	1.24E-05	Mann Whitney U
CH-PT	video	0.92	1.22	4.55E-10	Mann Whitney U
PT-CH	audio	1.39	2.54	8.13E-10	Mann Whitney U
PT-CH	video	1.34	2.07	6.20E-09	Mann Whitney U

After that, we analyzed the duration of FP-output, that is, the duration of full silent pauses that concur at the interpretation of the speech gestured segments. As shown in Table 6, the duration of these pauses produced by both control and experimental groups of trainee-interpreters in both directions are above the average level of pause duration registered during their whole interpreting performance, although the difference is not statistically significant, with p-value exceeding 0.05.

Therefore, we consider that during the production of gesture-accompanied source speech, the participants may have experienced certain difficulties leading to the prolongation of the silent pauses during production, but we cannot reject the possibility that it may have been caused by certain abnormal data within the dataset. Furthermore, the p-values for the audio group are 0.065 and 0.080, while the p-values for the video group are 0.718 and 0.479. The higher the p-value is, the truer the null hypothesis becomes, since it posits no significant difference between the two data sets. We therefore consider the prolonged pauses during interpreting is marginally significant for the control group but not for the experimental group.

TABLE 6
COMPARISON OF MEAN OP-INPUT DURATION BETWEEN EXPERIMENTAL AND CONTROL GROUPS AT THE PERCEPTION STAGE

Direction	Input Condition	Duration for all FP	Duration for FP-output	p-value	Test used
CH-PT	audio	0.95	0.97	0.065	Mann Whitney U
CH-PT	video	0.92	0.90	0.718	Mann Whitney U
PT-CH	audio	1.39	1.55	0.080	Mann Whitney U
PT-CH	video	1.34	1.49	0.479	Mann Whitney U

To further explore the impact of the visualization of semantic gestures, we selected further the OP for analysis and continue to focus on two types of them: OP-input and OP-output. Also, we conducted a comparative analysis of the mean data between the test group with video input and the control group with only audio input. Furthermore, we analyzed the FP's duration of the two directions of SI.

We first compared the OP-input of the two groups. As shown in Table 7, the experimental group shows an advantage in both directions with shorter OP duration than that the control group. The results also suggest that in the direction from Portuguese to Chinese the experimental group shows a larger advantage in terms of OP duration (0.03s) than in the direction from Chinese to Portuguese (0.07s). It is understandable since the semantic gesture input helped the comprehension of source speech in Portuguese, which is the second language for the participants. On the contrary, the visual inputs of semantic gestures in Chinese speech result in split of attention distributed to gestures, possibly competing with production resources of the interpreter.

TABLE 7
COMPARISON OF MEAN OP-INPUT DURATION BETWEEN EXPERIMENTAL AND CONTROL GROUPS AT THE PERCEPTION STAGE

	direction of interpreting	mean pause duration with **audio input**	mean pause duration with **video input**	p-value	test used
OP-input	CH-PT	0.646143969	0.673553498	0.08533405	Mann Whitney U
OP-input	PT-CH	0.855604762	0.794993072	0.14264704	Mann Whitney U

Besides the mean pause difference shown in Table 7, though not statistically significant, the advantage shown by the experimental group at the time point of output of the respective content (see Table 8) provides further evidence that gestural visual input does result in shorter silence in the SI output.

TABLE 8
COMPARISON OF OP-OUTPUT DURATION BETWEEN AUDIO AND VIDEO INPUT GROUP AT THE PRODUCTION STAGE

gesture condition	direction of interpreting	OP-output for **audio input** group	OP-output for **video input** group	p-value	test used
OP-output	CH-PT	0.73	0.50	2.2779E-10	Mann Whitney U
OP-output	PT-CH	0.99	0.42	3.3723E-16	Mann Whitney U

As shown in Table 8, we examined the duration of the parts of silent pauses within the SI output of the content of gesture-accompanied source speech segments, and a significant longer silent pause duration was found in the audio group (0.17s longer for Chinese to Portuguese direction and 0.57s longer for Portuguese to Chinese direction). These results indicate that with the visual access to the semantic gestures of the source speech speaker, the cognitive load of the experimental group was lower than that of the control group, which is demonstrated by a significant shorter pause duration while interpreting the content of these source speech segments. In addition, we elaborated a boxplot (see Figure 5) to demonstrate the differences between the audio and video group. As shown in the plot, the video group has a significant shorter pause duration in comparison with the audio group.

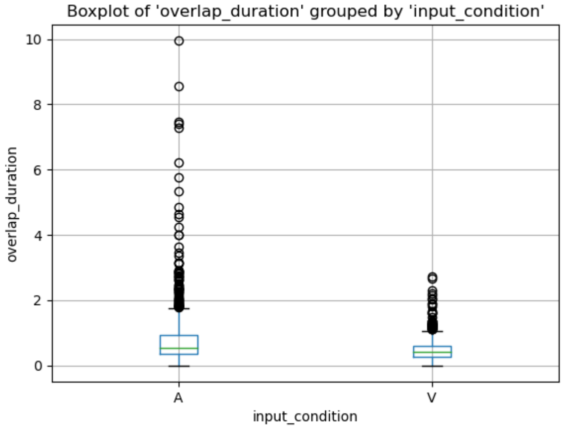

Figure 5. Boxplot for OP-Output Duration: Comparison Between the Video and Audio Input Groups

B. Comparison Between the Two Directions of SI

Apart from the gestures, we also executed an additional examination to the difference of full pauses in two distinct interpreting directions, that is, the Chinese to Portuguese direction and the Portuguese to Chinese direction. The results in Figure 6 and Table 10 showed that the overall pause duration for the retour direction, that is, from the L1 of interpreters (Chinese) to the L2 or interpreters (Portuguese) of the participants, is significantly shorter.

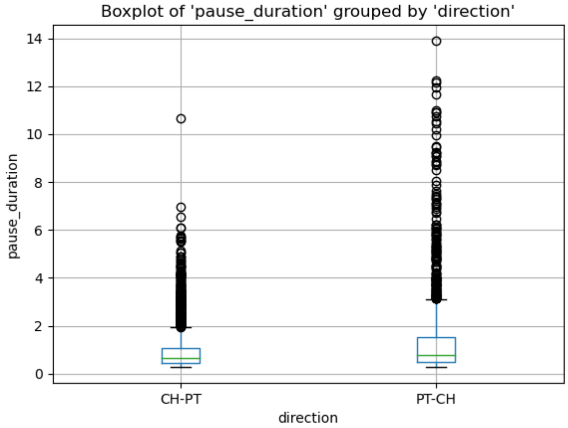

Figure 6. Boxplot for FP Duration: Comparison Between Chinese to Portuguese and Portuguese to Chinese Groups

TABLE 9
DIFFERENCE ON PAUSE DURATION BETWEEN VIDEO AND AUDIO INPUT GROUP

	FP duration for PT-CH	U statistic	p-value
0.93s	1.36s	2119887.00	0.0000

However, despite the overall shorter pause for the Chinese to Portuguese direction, the OP-output of the video input group, as previously demonstrated in Table 8, is shorter in the PT to CH direction than vice-verse. The results imply that although participants achieve better fluency in CH to PT direction overall, when it comes to the SI performance regarding the gestured segments, the experimental group achieves better fluency in PT to CH direction. A possible reason is that the cognitive resource demanded by content retention and lexical retrieval was mitigated due to the previous gesture visual input.

V. FURTHER DISCUSSION

We are requesting that you follow these guidelines as closely as possible. In this section, we present a further discussion apart from the data analysis, with post hoc analysis to explore how the silent pause duration is influenced by different input conditions and different direction of interpreting.

We conducted an integral analysis of RF regression (see Figure 7) to examine how the dependent variables - the duration of silent pauses (FP) - varies in response to changes in several factors (independent variables, or feature, as designated in RF regression model). These factors or features refer to the type of input condition (either video or audio) (coded as ˆinput_condition_A´ and ˆinput_condition_V´), the direction of interpretation (coded as ´direction_CH-PT´

and ʻdirection_PT-CHʼ), and the gesture-related condition (with FP represented by ʻTable_Table1´, FP-input represented by ʻTable_Table2_obs ´and FP-output represented by ʻTable_Table2_int´in the figures below). All the features mentioned above are analyzed by the RF regression model and their importance is shown in Figure 7.

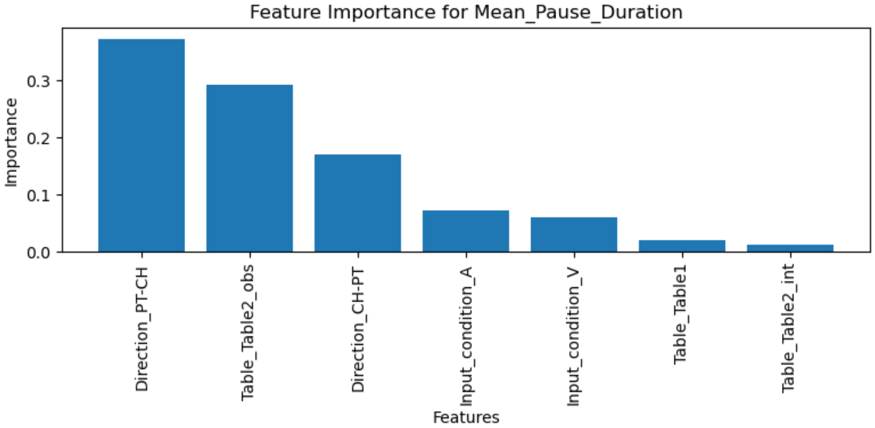

Figure 7. Feature Importance Analysis for Mean Pause Duration

The analysis of feature importance in Figure 7 shows that among all the features examined, the direction and the occurrence of gestures made by the source speech speaker (coded as ʻTable featuresʼ) have the most evident influence on the pause duration, while the input condition is of low importance when it comes to the prediction of the duration of pauses overall.

The partial dependence plots of Figures 8 (a), (b) and (c) show the extent to which different features influenced the prediction of the duration of FP. The x-axis in the PDP (partial dependent plots), ranging from zero to one, represents the non-occurrence or occurrence of a feature, such as ʻDirection_PT-CH´ in Figure 8 (a), with non-occurrence represented by 0 and occurrence by 1. The y-axis depicts changes in pause duration.

As shown in the PDP plots, the directions (coded as ʻDirection_PT-CH´and ʻDirection_CH-PTʼ) and the input of gesture-accompanied source-speech segments (coded as ʻTable_Table2_obsʼ) have caused a variance of around 0.5s on the mean duration of silent pauses.

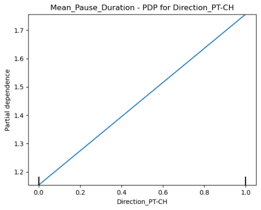

Figure 8 (a). PDP of PT to CH Feature for Integral Analysis of FP Duration for All Features.

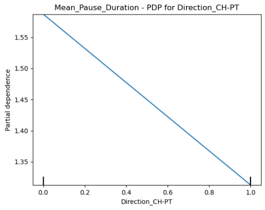

Figure 8 (b). PDP of CH to PT Feature for Integral Analysis of FP Duration for All Features.

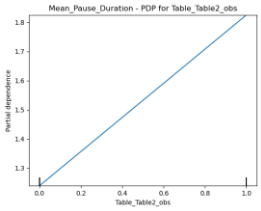

Figure 8 (c). PDP of Gesture Input Feature for Integral Analysis of FP Duration for All Features.

Figures 9 (a) and 9 (b) demonstrate the impact of the input condition on the duration of silent pauses in interpretation. It can be observed that as the audio input approaches 1, the duration of silent pauses tends to increase. Conversely, as the video input approaches 1, the duration of pauses tends to decrease. This suggests that when participants have visual input, the duration of pauses tends to become shorter. However, as we have mentioned in the analysis of feature importance in Figure 7, the input condition feature is of low importance for the RF regression model and the overall pause duration differences between the video group of interpreters and the audio group are not statistically significant (Table 2), we therefore consider relatively less relevant this variance caused by the video access to the source-speech speaker. Our results coincide with the earlier studies on the impact of visual input on the cognitive process of SI between English and German, as referred by Rennert, “for the most part, visual input appeared to have no appreciable positive or negative effect” (Rennert, 2008, p. 218). Early study about the impact of visual access on the performance of intelligibility and informativeness of professional interpreters conducted by Anderson (1994) also reported no statistical differences with the availability of visual input. At such, we consider that the visual access to the speaker may represent information highly automated for the interpreter, as it is identical to the visual information for daily audiovisual speech comprehension and demands no extra cognitive effort from the trainee interpreters (Wickens, 2002, p. 165). Indeed, it is our contention that a comprehensive investigation is required to determine whether visible speech relies on focal, ambient, or a combination of both visual resources. This examination can be conducted by employing physiological measures, such as eye tracking devices.

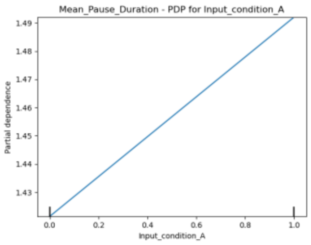

Figure 9 (a). PDP of Audio Input Condition Feature for Integral Analysis of FP Duration for All Features.

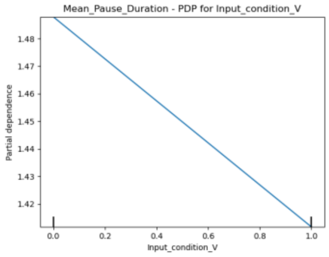

Figure 9 (b). PDP of Video Input Condition Feature for Integral Analysis of FP Duration for All Features.

Figure 10 (a) and (b) demonstrates the variation of OP (both OP-input and OP-output) duration caused by different input conditions. For the audio input group, the duration of OP increases, while for the video group, the OP duration decreases for about 0.125 seconds. In other words, when considering solely the SI output of the gestured segments in the source speech and the SI output during the input of these speech segments, the experimental group is expected to exhibit an average reduction of 0.125 seconds in silent intervals.

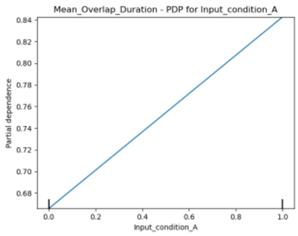

Figure 10 (a). PDP of Audio Input Feature for Integral RF Analysis of OP Duration for All Features.

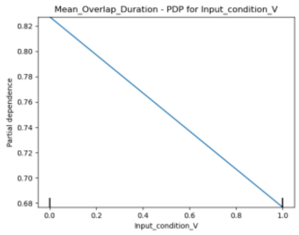

Figure 10 (b). PDP of Video Input Feature for Integral RF Analysis of OP Duration for All Features.

As for the gesture feature, we found that for the analysis of all the SI output audios, the overall OP duration is shorter during the interpretation of the speech gestured segments (code ‵Table_Table2_int ´in Figure 11 (a)). When we further conduct an RF analysis separately for the experimental and the control group, we found that the OP-output duration is only shorter for the experimental group (Figure 11 (c)), and for the control group (Figure 11 (b)), the situation is the contrary. That is to say, the control group's silence pause is actually longer when the interpreters translate the gestured segments of the source speech.

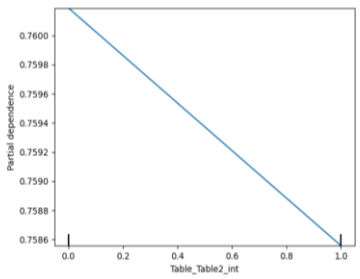

Figure 11 (a). PDP of OP-Output Feature in Integral RF Analysis of OP Duration for All Features.

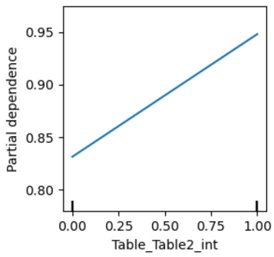

Figure 11 (b). PDP of OP-Output Feature in RF Analysis of OP Duration for the Audio Input Group.

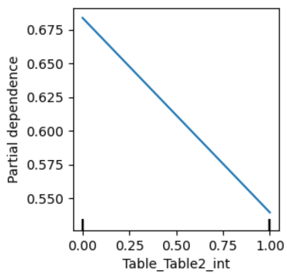

Figure 11 (c). PDP of OP-Output Feature in RF Analysis of OP Duration for the Video Input Group.

Regarding the direction of interpreting, we found a longer pauses on the direction from PT to CH (Figure 12 (a)), that is, longer pauses from second language (L2) to first language (L1) of the participants. This finding contradicts to some of the previous studies, which argues the longer pauses in the direction from first (L1) to second language (L2) of interpreter (Gumul, 2021; Lin et al., 2018). This unexpected finding might be related with the selection of source speeches. As for the two authentical Portuguese and Chinese speeches that we selected for experiment, the Chinese speech has relatively higher semantically related gestures (34 for Portuguese speech versus 36 for Chinese speech) and more beat gestures (36 for Portuguese speech versus 48 for Chinese speech). At such, we expect a future examination into the influence of directionality on the cognitive load of SI task and its interaction with the impact of various types of visual inputs available for the interpreter with video access to the source-speech speaker.

As previously mentioned in Table 3, a separate RF analysis was conducted to investigate the impact of the directionality on the duration of different types of pauses including OP-output, using data extracted separately from CH to PT and PT to CH interpretation audios. Figure 20 illustrates the modulation of OP-output duration by the video input feature in PT to CH direction while Figure 12 demonstrates its influence on OP-output duration in the CH to PT direction.

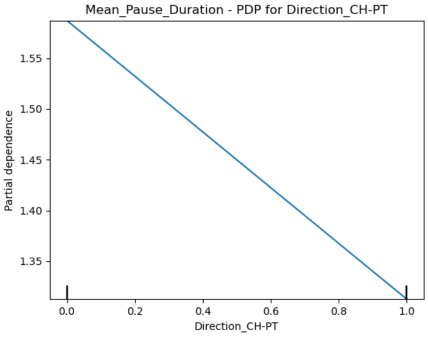

Figure 12 (a). PDP of CH-PT Feature for Integral RF Analysis of OP Duration for All Features.

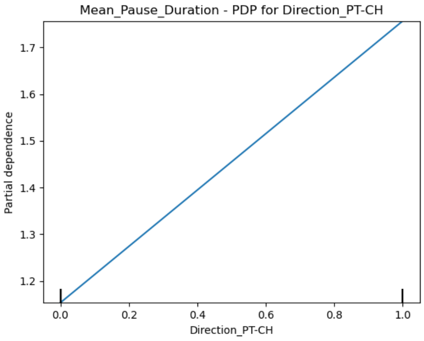

Figure 12 (b). PDP of PT-CH Feature for Integral RF Analysis of OP Duration for All Features.

We also found the facilitative effect of multimodal input to be more conspicuous in the direction from Language 2 to 1 (approximately 0.25 second) than From 1 to 2 (approximately 0.4 second) as shown in Figures 13 (a) and (b). The finding corroborates the beneficial effect of semantic gestures on the comprehension of SI second language information revealed in previous study conducted by Arbona et al. (2023). We consider that while rendering from Portuguese, the redundant verbal information provided by semantic gestures turns out to be more crucial for the comprehension and retention of Portuguese source speech than for the mother tongue of the interpreters.

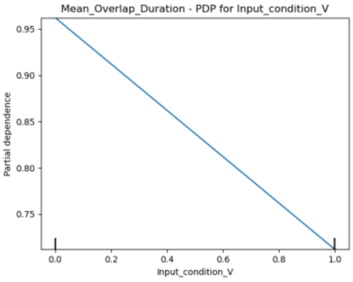

Figure 13 (a). PDP of Video Input Feature in RF Analysis of OP Duration for PT to CH Direction.

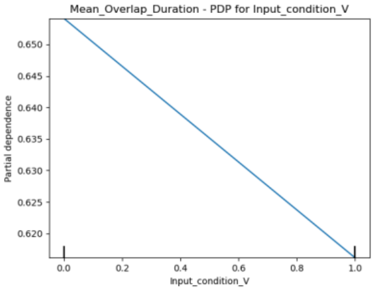

Figure 13 (b). PDP of Video Input Feature in RF Analysis of OP Duration for CH to PT Direction.

VI. Conclusion

In the present study, we examined the cognitive influence of visual access to the source speech speaker on trainee interpreters between Chinese and Portuguese, by delving into the silent pause durations in SI output. For the purpose, we examined the duration of full silent pauses (FP), the overlapped pauses (OP) concurring at the speech gestured segments/input (OP-input) and concurring at the interpretation of the speech gestured segments/output (OP-output), through an experiment of trainee interpreters.

Through the experiment, we found that both experimental group (with video) and control group (with only audio) experienced longer silent pause around the moment of the gestured segments of the source speech, which indicates the source speech segments accompanied by semantic gestures may represent one of the most cognitively demanding part of the entire speech. Therefore, at the perception and cognition stage (Seeber, 2017) of these segments, i.e., when hearing or seeing the gestures source speech segments, both the experimental and control groups of trainee interpreters shows a longer silent pause (FP-input) than their average level.

Although the results showed no significant differences between the full pause duration of video and audio input groups, participants with visual access to the source speech speaker have significantly lower OP-output duration in comparison with the audio input group in the interpretation of the critical gestured segments of the source speech. Moreover, the gap is larger in the direction from L2 to L1. Therefore, we draw a preliminary conclusion that the visualization of semantic gestures may have contributed to better comprehension and retention of information of these critical segments, which has positive effect by resulting in shorter pauses in the interpretation performance. We believe, for the experimental group, the speech segments accompanied by semantic gestures are coded not only verbally but also spatially due to the presence of the gestures produced by the source speech speaker, can result in better memory performance (Bonnici et al., 2016, p. 5466) and therefore alleviated cognitive load during the SI output of the respective segments.

Also, there are some limitations for present study. When comparing the differences of pause duration for the speech gestured segments between the experimental and control groups, we could not exclude the possible influence of other visual cues received by the participants with visual access, such as facial expressions, lip movements, etc. Nevertheless, we believe that the no significant difference on the overall pause durations between the two groups partially eliminates the possibility of the results being disturbed by those visual cues.

Finally, we consider the present study as a good tentative for exploration of the impact of semantic gesture inputs on the cognitive fluency of the interpreters, which also calls for future studies about other factors with potential influence on the results, such as the lip movements or the facial expressions of the source speech speaker, to further examine human cognition and its synergetic factors or constraints in interpreting tasks.

Acknowledgements

This work was supported by a Macao Polytechnic University Research Project (RP/FLT-11/2022).

References

[1] Anderson, L. (1979). *Simultaneous interpretation: Contextual and translation aspects*. https://doi.org/https://doi.org/10.1075/btl.3.11

[2] Anderson, L. (1994). Simultaneous interpretation: Contextual and translation aspects. InS. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. https://doi.org/https://doi.org/10.1075/btl.3.11

[3] Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences, 4*(11), 417-423. https://doi.org/https://doi.org/10.1016/S1364-6613(00)01538-2

[4] Balzani, M. (1990). *le contact visual en interpretation simultanée*. Paper presented at the Aspects of Applied and Experimental Research on Conference Interpretation.

[5] Baxter, R. N. (2016). Exploring the possible effects of visual presentations on synchronicity and lag in simultaneous interpreting. *Sendebar*, *27*, 9-23. Retrieved June 30, 2023, from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwje-qS7loyEAxVqLEQIHfhpD8YQFnoECA4QAQ&url=https%3A%2F%2Frevistaseug.ugr.es%2Findex.php%2Fsendebar%2Farticle%2Fdownload%2F3834%2F5052%2F11845&usg=AOvVaw0hu8OPaxB5gXWfGH8AP-YJ&opi=89978449

[6] Bonnici, H. M., Richter, F. R., Yazar, Y., & Simons, J. S. (2016). Multimodal feature integration in the angular gyrus during episodic and semantic retrieval. *Journal of Neuroscience*, *36*(20), 5462-5471. https://doi.org/10.1523/JNEUROSCI.4310-15.2016

[7] Gieshoff, A. C. (2017). Audiovisual speech decreases the number of cognate translations in simultaneous interpreting. *Empirical modelling of translation and interpreting*, *7*, 313. https://doi.org/10.5281/zenodo.1090978

[8] Gieshoff, A. C. (2018). *The impact of audio-visual speech on work-load in simultaneous interpreting.* (Doctor Dissertation). Johannes Gutenberg-Universität Mainz, Mainz. JGU-Publikationen database.

[9] Gieshoff, A. C. (2021a). Does it help to see the speaker's lip movements? An investigation of cognitive load and mental effort in simultaneous interpreting. *Translation, Cognition & Behavior*. https://doi.org/10.1075/tcb.00049.gie

[10] Gieshoff, A. C. (2021b). The impact of visible lip movements on silent pauses in simultaneous interpreting. *Interpreting*. https://doi.org/https://doi.org/10.1075/intp.00061.gie

[11] Goldman-Eisler, F. (1958). The predictability of words in context and the length of pauses in speech. *Language and Speech, 1*(3), 226-231. https://doi.org/https://doi.org/10.1177/002383095800100308

[12] Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*, *63*(4), 308-319. https://doi.org/https://doi.org/10.1198/tast.2009.08199

[13] Grosjean, F., & Deschamps, A. (1972). Analyse des variables temporelles du français spontané. *Phonetica*, *26*(3), 129-156. https://doi.org/https://doi.org/10.1159/000259407

[14] Han, L., Lu, J., Wen, Z. E., & Tian, Y. (2023). *Momentary engagement in simultaneous versus consecutive interpreting: through the lens of translanguaging and CDST.* https://doi.org/https://doi.org/10.3389/fpsyg.2023.1180379

[15] Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with" articulatory" pauses. *Language and Speech*, *26*(3), 203-214. https://doi.org/https://doi.org/10.1177/002383098302600302

[16] Jiang, X. L., & Jiang, Y. (2020). Effect of dependency distance of source text on disfluencies in interpreting. *Lingua*, *243*, 18. https://doi.org/https://doi.org/10.1016/j.lingua.2020.102873

[17] Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809-854. https://doi.org/10.1111/lang.12084

[18] Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta psychologica*, *134*(3), 372-384. https://doi.org/10.1016/j.actpsy.2010.03.010

[19] Lin, Y., Lv, Q., & Liang, J. (2018). Predicting fluency with language proficiency, working memory, and directionality in simultaneous interpreting. *Frontiers in psychology*, *9*(1), 1543. https://doi.org/10.1016/j.plrev.2018.06

[20] McKnight, P. E., & Najab, J. (2010). Mann‐Whitney U Test. *The Corsini encyclopedia of psychology*, 1-1. https://doi.org/https://doi.org/10.1002/9780470479216.corpsy0524

[21] Mead, P. (2000). Control of pauses by trainee interpreters in their A and B languages. *The interpreters' newsletter*, *10*(200), 89-102. Retrieved June 30, 2023, from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1a2485651848b1baeaed9cf12044976064d3a2a3

[22] Mikkelson, H., & Jourdenais, R. (2015). *The Routledge handbook of interpreting* (H. Mikkelson & R. Jourdenais Eds.). London: Routledge New York.

[23] Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247-279. https://doi.org/10.1016/0010-0285(82)90010-x

[24] Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., . . . Foxe, J. J. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of neurophysiology*, *96*(2), 721-729. https://doi.org/10.1152/jn.00285.2006

[25] Moser-Mercer, B. (2005). Remote interpreting: The crucial role of presence. *Bulletin vals-asla*, *81*, 73-97. https://doi.org/https://benjamins.com/online/hts/articles/rem1/1000?filter-language=uk

[26] Plevoets, K., & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting: A corpus-based regression analysis. Translation and Interpreting Studies. *The Journal of the American Translation and Interpreting Studies Association*, *11*(2), 202-224. https://doi.org/10.1075/tis.11.2.04ple

[27] Prandi, B. (2023). *Computer-assisted simultaneous interpreting: A cognitive-experimental study on terminology* (Vol. 22): Language Science Press.

[28] Raupach, M. (2011). Temporal variables in first and second language speech production. In *Temporal variables in speech* (pp. 263-270): De Gruyter Mouton.

[29] Rennert, S. (2008). Visual input in simultaneous interpreting. *Meta: journal des traducteurs/Meta: Translators' Journal*, *53*(1), 204-217. https://doi.org/https://doi.org/10.7202/017983ar

[30] Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes*, *14*(4), 423-441. https://doi.org/https://doi.org/10.1080/01638539109544795

[31] Seeber, K. G. (2007). *Thinking outside the cube: Modeling language processing tasks in a multiple resource paradigm.* Paper presented at the Eighth Annual Conference of the International Speech Communication Association.

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

[33] Seeber, K. G. (2012). *Multimodal input in Simultaneous Interpreting: An eye-tracking experiment.* Paper presented at the Proceedings of the 1st International Conference TRANSLATA, Translation & Interpreting Research: yesterday-today-tomorrow.

[34] Seeber, K. G. (2017). Multimodal processing in simultaneous interpreting. *The handbook of translation and cognition*, 461-475. https://doi.org/https://doi.org/10.1002/9781119241485.ch25

[35] Shreve, G. M., Lacruz, I., & Angelone, E. (2011). Sight translation and speech disfluency Performance analysis as a window to cognitive translation processes. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies* (Vol. 94, pp. 93-120).

[36] Skehan, P. (2003). Task-based instruction. *Language teaching*, *36*(1), 1-14. https://doi.org/10.5296/ijl.v4i3.2203

[37] Song, S. (2020). *Fluency in simultaneous interpreting of trainee interpreters: the perspectives of cognitive, utterance and perceived fluency.* (Ph.D.). Hong Kong Polytechnic University, Retrieved June 30, 2023, from https://theses.lib.polyu.edu.hk/bitstream/200/10418/1/991022378657903411.pdf

[38] Song, S., & Li, D. (2020). The Predicting Power of Cognitive Fluency for the Development of Utterance Fluency in Simultaneous Interpreting. *Frontiers in psychology*, *11*, 1864. https://doi.org/https://doi.org/10.3389/fpsyg.2020.01864

[39] Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of America*, *26*(2), 212-215. https://doi.org/https://doi.org/10.1121/1.1907309

[40] Teder-Sälejärvi, W., McDonald, J., Di Russo, F., & Hillyard, S. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*(1), 106-114. https://doi.org/ 10.1016/s0926-6410(02)00065-4

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

[42] Wang, B., & Li, T. (2015). An empirical study of pauses in Chinese-English simultaneous interpreting. *Perspectives*, *23*(1), 124-142. Retrieved June 30, 2023, from https://www.polyu.edu.hk/edc/tdg/userfiles/file/42A5_paper2.pdf

[43] Wang, X., & Wang, B. (2022). Identifying fluency parameters for a machine-learning-based automated interpreting assessment system. *Perspectives*, 1-17. Retrieved June 30, 2023, from https://doi.org/10.1080/0907676X.2022.2133618

[44] Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177. https://doi.org/https://doi.org/10.1080/14639220210123806

[45] Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research.* Paper presented at the 5th international conference on language resources and evaluation (LREC 2006).

[46] Xin, S. (2020). *ying han jiao ti chuan yi zhong bu dang ting dun de chu xian yuan yin he ying dui ce lue* [Causes and coping strategies for the emergence of inappropriate pauses in English-Chinese consecutive interpreting]. (Unpublished master thesis). Shanghai International Studies University, Retrieved June 30, 2023, from 10.27316/d.cnki.gswyu.2020.000139

[47] Yang, S. (2019). *Investigating the Effect of Speech Rate on the Cognitive Load in Simultaneous Interpreting with Text.* University of Macau, Retrieved June 30, 2023, from https://www.proquest.com/openview/150fd38d1a83c44ecd1e6f43671e2221/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y

[48] Yang, S., Li, D., & Lei, V. L. C. (2020). The impact of source text presence on simultaneous interpreting performance in fast speeches: Will it help trainees or not? *Babel*, *66*(4-5), 588-603. https://doi.org/10.1075/babel.00189.yan

[49] Zhao, N. (2022). Speech Disfluencies in Consecutive Interpreting by Student Interpreters: The Role of Language Proficiency, Working Memory, and Anxiety. *Frontiers in psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.881778

Yunxiao Jiang is currently a second-year PhD student in Interpreting Studies at Macao Polytechnic University in China. Her primary areas of academic interest are meta-research on translation and interpreting, psycholinguistics, and interpreting studies. She is currently working on the impact of semantically related gestures occurred in source speech on the performance of interpreters. E-mail: sarrrinha@163.com.

Lili Han (Ph.D.) is Associate Professor of the Faculty of Applied Sciences of Macao Polytechnic University, Macau. Over the last decade, Dr. Han has lectured and conducted research in interpreting studies, acting as trainer for the Conference Interpreting (Chinese-Portuguese-English) course in partnership with the DG (SCIC) of the European Commission. Her research interests include interpreting studies, intercultural studies, language and translation policy studies, interpreting testing & assessment, and computer-aided interpreting. Email: hanlili@mpu.edu.mo.

Yuqi Sun, PhD in Linguistics from the Pontifical Catholic University of Rio Grande do Sul, is currently an assistant professor in the Department of Portuguese at the University of Macau. With extensive experience in the field of linguistics, Dr. Sun Yuqi's research focuses on pragmatics and interpreting studies. ORCID ID: https://orcid.org/0000-0002-7310-1385. E-mail: sunyuqi@um.edu.mo.