Bilingualism: Language and Cognition

cambridge.org/bil

Semantically related gestures facilitate language comprehension during simultaneous interpreting

Eléonore Arbona[1], Kilian G. Seeber[2] and Marianne Gullberg[3]

[1]Faculty of Translation and Interpreting, University of Geneva, Geneva, Switzerland; [2]Faculty of Translation and Interpreting, University of Geneva, Geneva, Switzerland and [3]Centre for Languages and Literature and Lund University Humanities Lab, Lund University, Lund, Sweden

Research Article

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

Received: 30 July 2022
Revised: 22 August 2022
Accepted: 22 August 2022
First published online: 10 October 2022

Keywords:
simultaneous interpreting; language comprehension; gestures; multimodality; bilingualism

Author for correspondence:
Eléonore Arbona,
FTI, Département d'Interprétation Université de Genève 40 bd du Pont-d'Arve
CH-1211 Genève 4
E-mail: Eleonore.Arbona@unige.ch

Abstract

Manual co-speech gestures can facilitate language comprehension, but do they influence language comprehension in simultaneous interpreters, and if so, is this influence modulated by simultaneous interpreting (SI) and/or by interpreting experience? In a picture-matching task, 24 professional interpreters and 24 professional translators were exposed to utterances accompanied by semantically matching representational gestures, semantically unrelated pragmatic gestures, or no gestures while viewing passively (interpreters and translators) or during SI (interpreters only). During passive viewing, both groups were faster with semantically related than with semantically unrelated gestures. During SI, interpreters showed the same result. The results suggest that language comprehension is sensitive to the semantic relationship between speech and gesture, and facilitated when speech and gestures are semantically linked. This sensitivity is not modulated by SI or interpreting experience. Thus, despite simultaneous interpreters' extreme language use, multimodal language processing facilitates comprehension in SI the same way as in all other language processing.

Introduction

Co-speech gestures are body movements that accompany speech, and are temporally, semantically and pragmatically coordinated with speech (Kendon, 2004; McNeill, 1985). Manual co-speech gestures (i.e., co-speech gestures made with the hands) facilitate spoken language comprehension in monolingual first language (L1) and second language (L2) settings (e.g., Dahl & Ludvigsen, 2014; Hostetter, 2011; Sueyoshi & Hardison, 2005) as they offer a parallel representation of meaning which can be redundant, supplementary, etc. Simultaneous interpreting (SI), an instance of extreme language use (Hervais-Adelman, Moser-Mercer, & Golestani, 2015), involves simultaneous processing and comprehension of spoken language input (Seeber, 2017) and production of language output in another spoken language[1]. Thus, one could expect simultaneous interpreters to benefit from co-speech gestures during language comprehension just like L1 and L2 speakers do. However, the fact that interpreters also produce verbal output while comprehending may modulate the effect of gesture on language comprehension. Yet there is some evidence suggesting that interpreting expertise might positively influence cognitive performance, e.g., dual-task performance (Strobach, Becker, Schubert, & Kühn, 2015) or cognitive flexibility (Yudes, Macizo, & Bajo, 2011), suggesting that interpreters might be better than other bilinguals at attending to visual and auditory input in parallel. However, empirical data are scarce on the potential influence of gestures on language comprehension in SI.

In this study we therefore strive to bridge this gap, and explore whether gestures influence language comprehension during SI. Specifically, we look at simultaneous interpreters' comprehension of semantically related/unrelated co-speech manual gestures during SI and during passive viewing/listening, in comparison to a bilingual group with no interpreting experience.

The role of gestures in language comprehension

A growing body of work shows that co-speech gestures and speech are intimately related, forming a so-called integrated system (McNeill, 1985, 1992, 2005). Indeed, co-speech gestures and speech have been shown to develop, break down, and be processed in parallel (Capirci & Volterra, 2008; Colletta, Guidetti, Capirci, Cristilli, Demir, Kunene-Nicolas, & Levine, 2015; Goldin-Meadow, 2003; Graziano & Gullberg, 2018; Holle & Gunter, 2007; Kelly, Özyürek, & Maris, 2009; Mayberry & Jaques, 2000; Mayberry & Nicoladis, 2000; Rose, 2006; Wu &

[1]Please note that we only investigate spoken language interpreting. While sign language interpreting is also a form of simultaneous interpreting, it is beyond the scope of this paper.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Coulson, 2007; for a review of the integrated relationship between gesture and speech, see Kelly, 2017).

Gestures can represent the properties of entities and events talked about drawing on iconicity or similarity of shape, size, movement (McNeill, 1992); e.g., when a speaker makes a large, circular gesture while saying, "It was a big, round one" (Church, Garber, & Rogalski, 2007, p. 138). Such representational gestures typically express information that is semantically related to concurrent speech. Other gestures can express pragmatic aspects of speech such as rhythm, speech acts, stance, or aspects of discourse structure (Kendon, 1995, 2004), such as when a speaker rotates both forearms outwards with extended fingers to a "palm up" position to display incapacity, powerlessness or indifference (Debras, 2017). Such pragmatic gestures have a more complex semantic relationship to concurrent speech compared to representational gestures. There is considerable evidence that semantically related (representational) gestures facilitate spoken language comprehension in naïve listeners when the meaning in speech and gesture is congruent (see Hostetter, 2011 for a review). When processing multimodal information, comprehenders build a single unified meaning representation without necessarily realising which particular channel the information came from (Cassell, McNeill, & McCullough, 1999; Gullberg & Kita, 2009). Furthermore, priming studies using multimodal stimuli have revealed an interaction between semantically related gestures and speech, even when one modality is irrelevant to the experimental task (Kelly, Creigh, & Bartolotti, 2010; Kelly, Healey, Özyürek, & Holler, 2015; Langton & Bruce, 2000). The integrated-systems hypothesis, which is based on these observations, posits that gestures necessarily influence the processing of speech, while speech necessarily influences the processing of gestures (Kelly et al., 2009). Furthermore, empirical evidence suggests that co-speech gestures may also contribute to language comprehension in L2 speakers (Dahl & Ludvigsen, 2014; Sueyoshi & Hardison, 2005). Although the link between speech and gestures is thus undisputed (Gullberg, 2006), speech-gesture integration during a complex task such as SI remains under-explored.

Multimodality in SI

SI is considered a complex process (Frauenfelder & Schriefers, 1997; Moser-Mercer, 2000) since it combines concurrent spoken language comprehension and production in two distinct languages. From a cognitive point of view, the language comprehension component in SI is likely to share common features with other language comprehension tasks (Seeber, 2017). For example, just like during ordinary language comprehension, interpreters process speakers' input while having access to various sources of visual information, including gestures (Galvão, 2013; Gieshoff, 2018; Seeber, 2017). Indeed, practitioners generally deem visual access necessary for successful interpretation (Bühler, 1985). More specifically, hand gestures and facial expressions are considered to be the most important visual sources of information, since they are viewed as facilitating understanding and emphasis (Rennert, 2008). What is more, visual access to speakers, including to their gestures, is enshrined in the working conditions issued by the International Association of Conference Interpreters (AIIC, 2007) and in ISO standards (ISO, 2016b, 2017). More generally, some have argued that interpreters should use any piece of information that can make language comprehension easier or faster, since comprehension is a key component of SI (Bühler, 1985). Moreover, there is some evidence suggesting that interpreting expertise might positively influence cognitive performance, e.g., dual-task performance (Strobach et al., 2015) or cognitive flexibility (Yudes et al., 2011), suggesting that interpreters might be better than other bilinguals at attending to visual and auditory input in parallel. Thus, even when engaging in SI, interpreters might be able to benefit from gestures just like other comprehenders do in L1 and L2 settings. To date, however, this assumption has not been empirically corroborated.

Until recently, the focus of translation and interpreting studies has been on the investigation of written and oral texts as verbal artifacts, meaning that written and spoken discourse has been studied in isolation from other non-verbal resources (González, 2014). This is reflected in many influential models of SI that focus on the verbal signal and have failed to give sufficient prominence to the integration of different channels. Some efforts have been made to document and empirically measure the impact of visual access to speakers on SI (Anderson, 1994; Bacigalupe, 1999; Balzani, 1990; Rennert, 2008; Tommola & Lindholm, 1995). For example, Anderson (1994) carried out an experiment with twelve professional interpreters to assess the effect of visual access on SI, and found no significant difference between the audiovisual and the audio-only condition. Tommola and Lindholm (1995) used a similar setup with eight experienced interpreters, and found no significant effect of visual access either. However, the set-ups did not allow us to ascertain whether interpreters were attending to the visual stimuli in the audiovisual condition and, if they were, what cues they might have processed. Therefore, it remains unclear how interpreters actually allocate visual attention to (and how they process) specific visual cues, especially gestures.

Investigating multimodal processing in SI using eye-tracking

In many experimental paradigms, eye-tracking techniques allow for the minimally-invasive recording of perceivers' visual behaviour toward gestures without compromising the ecological validity of the task (Gullberg & Holmqvist, 1999). In gesture studies, eye-tracking has been used in experiments focusing notably on the perception and processing of gestural information (Beattie, Webster, & Ross, 2010; Gullberg & Holmqvist, 1999, 2006; Gullberg & Kita, 2009) as well as on the integration of gesture and speech during reference resolution (Campana, Silverman, Tanenhaus, Bennetto, & Packard, 2005). Gullberg and Holmqvist (1999, 2006) established that, both in live and video conditions, addressees looked at the speaker's face the vast majority of the time while gestures were mainly perceived through peripheral vision; that said, gestural holds (a momentary cessation of movement in a gesture) and gestures that speakers themselves looked at attracted addressees' fixations more frequently. Beattie et al. (2010) found that short character-viewpoint gestures attracted more fixations than other gestures, suggesting that they are particularly information-rich.

As to interpreting studies, several studies have used eye-tracking to examine SI as a multimodal rather than a purely verbal process (Galvão & Rodrigues, 2010; Seeber, 2017; Stachowiak-Szymczak, 2019). In an eye-tracking experiment using pictures rather than gestures, Stachowiak-Szymczak (2019) found that visual and auditory input were integrated in SI, highlighting the multimodal nature of the task. Seeber (2011) conducted an eye-tracking experiment relating interpreters' fixations on visual information to the auditory content, concluding that interpreters do attend to visual cues, including gestured numbers.

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

To summarise, gestures and speech form an integrated system where both channels influence each other, and gestures have been shown to facilitate language comprehension in native and non-native speakers alike. While implying concurrent language comprehension and production in distinct languages, SI comprises a language comprehension component. If semantically related gestures have the potential to facilitate language comprehension during an extreme language task such as SI, interpreters should be expected to benefit from access to such gestures. The studies on visual access in SI conducted up until now have presented concomitant visual cues, not just gestures, and/or have not allowed for the establishment of a direct link between visual cues and language comprehension. Thus, it remains unclear whether gestures have the same influence during SI. Yet, the majority of interpreting practitioners, the main professional association and a growing number of scholars recognise the multimodal character of SI, including the potential positive influence of gestures. Moreover, evidence suggests that interpreters might be better than other bilinguals at attending to visual and auditory input in parallel.

The current study

This study aimed to investigate the potential facilitatory effects of semantically related gestures on simultaneous interpreters' language comprehension. The first experiment looked at task-contingent differences in processing audiovisual signals. It compared how interpreters comprehend audiovisual signals during SI versus during passive viewing/listening as measured by a picture matching task, and examined the effect of semantically related gestures (target gesture condition), semantically unrelated gestures (control gesture condition), and the absence of gestures (no-gesture condition) on comprehension. We expected that a congruent speech-gesture meaning pair (semantically related gestures) would be more easily processed than one where the speech-gesture meaning relationship is less clear (semantically unrelated gestures). We thus hypothesised a facilitatory influence of semantically congruent gestures on language comprehension. Language comprehension was measured through response accuracy and reaction times in the picture-matching task. Using eye-tracking, we also measured overt visual attention to gestures, operationalised as total visual dwell time (the total duration of fixations on a particular area of interest) on the speaker's gesture space, a pre-defined area of interest in front of the speaker going from the speaker's shoulders to her hips, since this is where speakers usually gesture (McNeill, 1992, p. 86). Monitoring what participants look at during the stroke, the meaningful part of the gesture, enabled us to examine the extent to which overt visual attention to gestures correlates with response accuracy and reaction times. The experiment aimed to address the following questions:

1. Do simultaneous interpreters integrate gestural information during language comprehension? 2. If so, is such integration affected by task (cf. SI versus during passive viewing/listening)? 3. Do simultaneous interpreters visually attend to gestures?

The second experiment examined experience-contingent differences in processing audiovisual signals. It compared language comprehension during passive listening/viewing in two groups: an experimental group of professional simultaneous interpreters and a comparison group of professional translators (i.e., language professionals who change written words into another written language; the main difference with simultaneous interpreters being that they work with text rather than with speech in realtime and that there is no simultaneity requirement) without SI experience. The aim was to determine whether interpreters behave differently than other bilinguals due to their SI experience, which could have influenced their performance in the first experiment. The same variables were analysed as in the first experiment to address the following questions:

4. Do translators integrate gestural information during language comprehension in the same way as interpreters? 5. Do translators visually attend to gestures in the same way as interpreters?

Table 1. Background information provided in the language background questionnaire.

	Interpreters ($n = 24$)	
	M	*SD*
Background		
Age *(yrs)*	47.38	11.37
Languages spoken	4.75	1.45
Professional experience *(yrs)*	19.25	11.00
French language		
Age of Acquisition *(yrs)*	2.65	3.22
Age of Fluency *(yrs)*	4.58	3.39
Duration of exposure[2] *(yrs)*	39.87	12.42
Current exposure *(%)*	54.95 *(n = 22)*	16.07
English language		
Age of Acquisition *(yrs)*	8.77	4.44
Age of Fluency *(yrs)*	17.42	6.83
Duration of exposure[3] *(yrs)*	2.46	3.42
Current exposure *(%)*	20.86 *(n = 22)*	8.23
Self-rated English proficiency		
Speaking	7.83	1.49
Reading	9.17	0.70
Listening	8.88	0.74

First experiment

Method

Participants

Twenty-four professional conference interpreters participated in the study[4] (see Table 1). They were recruited via e-mail describing the eligibility criteria. Participants completed an adapted version of the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007). All participants had normal or corrected-to-normal vision and reported no

[2]Number of years participants spent in a country or region where the relevant language is spoken

[3]See above

[4]One additional participant could not be tested as calibration of the eye-tracker could not be performed.

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

language disorders. Participants' L1 was French (A language[5]), their L2 English (A, B or C language[6]). Twenty-one of the 24 participants were either members of the International Association of Conference Interpreters (AIIC), or accredited by international organisations such as the United Nations, or both. The three remaining participants were professional conference interpreters based in Geneva.

All participants gave written informed consent. The experiment was approved by the Faculty of Translation and Interpreting's Ethics Committee. No participant was involved in the norming of the stimuli.

Task and materials

Participants were asked to either simultaneously interpret (SI activity) or to watch (passive viewing/listening activity) short video clips of a speaker uttering two sentences (e.g., "Look at the terrace! Last Monday, the girl picked the lemon"). The second sentence was either accompanied by a semantically related gesture, a semantically unrelated gesture, or no gesture. Participants were then presented with two drawings corresponding to an action verb, one target drawing (e.g., picking a lemon) and one distractor (e.g., squeezing a lemon), and asked to choose the drawing corresponding to the video by pressing a button. There was no time limit. This picture-matching task was used to probe language comprehension. Accuracy and response times were recorded. Since they express meaning differently from both speech and gesture, drawings enabled us to implicitly probe gesture content.

Speech

We created a first set of 30 utterances following one of two patterns: adverbial phrase of time, agent, verb and patient (e.g., "Last Monday, the girl *picked* the lemon."), or adverbial phrase of time, agent, verb, preposition and indication of location (e.g., "Two weeks ago, the boy *swung on* the rope"). The target word was the main verb. A short introductory sentence (e.g., "Look at the terrace") was added to ensure interpreters would interpret simultaneously.

We then created a second set of 30 sentences replacing the verbs with equally plausible candidates (e.g., "Last Monday, the girl *squeezed* the lemon." or "Two weeks ago, the boy *climbed up* the rope"). The resulting 60 sentences (word count: $M = 11.6$, $SD = 0.8$) were assigned to two matched stimulus lists. Target verb frequency indications were obtained from the Corpus of Contemporary American English (Davies, 2008) and two lists of sentences were created to balance verb frequency. Mean verb frequency was 72,788 ($SD = 79{,}898$) in List A and 62,041 ($SD = 74{,}694$) in List B, with no significant difference across lists, $p > .6$. Sentence plausibility was rated separately by 28 French and 28 English speakers (n raters = 56) on 6-point Likert-type scales (from 1, "very implausible" to 6, "very plausible"). Mean sentence plausibility (the average of the French and the English rating) was 3 ($SD = 0.9$) in List A and 3 ($SD = 0.9$) in List B, with no significant difference across lists, $p > .8$. Raters who rated plausibility did not participate in the norming of pictures.

Gestures

We devised manual gestures to accompany the sentences in the semantically related gesture condition versus in the semantically unrelated gesture condition. Semantically related gestures were representational gestures corresponding to the content of the target verb. For example, for "squeeze the lemon", the speaker performed a squeezing gesture: "right hand: hand half open in front of speaker, palm facing down, then fist closing with a rotation of the wrist" (see Figure 1a). Semantically related gestures depicted path rather than manner of movement in motion verbs (i.e., they showed a trajectory, e.g., *going up*, but did not provide information about manner of motion, e.g., no wiggling of fingers to indicate climbing).

Semantically unrelated gestures were instantiated as pragmatic gestures (Kendon, 2004) with no semantic relationship to the target verb. We used five forms: the Open Hand Prone and Open Hand Supine families (Kendon, 2004, pp. 248–283), the 'slice gesture', and the 'power grip' (Streeck, 2008), and the 'flick of the hand' (McNeill, 1992, p. 16). For example, for "squeeze the lemon", the speaker performed the gesture called "Open hand prone ('palm down') – vertical palm". Here the speaker's palm and forearm are vertical so that the palm of the hand faces directly away from the speaker (Fig. 1d).

All sentences were recorded audiovisually by a right-handed female speaker of North-American English in a sound-proof recording studio in controlled lighting conditions. Three versions of each sentence pair were recorded: one in which the speaker did not gesture while uttering the sentences (no-gesture condition), one in which the speaker performed a pragmatic hand gesture while uttering the target verb (semantically unrelated gesture condition), and one in which she performed a representational hand gesture while uttering the target verb (semantically related gesture condition). Sentences were read from a prompter. The general intended gestural movement for each clip was described to the speaker but she was asked to perform her own version of them so that they would be as natural as possible. All gestures were performed with the speaker's dominant (right) hand. The mean duration of the audiovisual recordings was 4.8 seconds ($SD = 0.4$, range 3.7–5.5). Horizontally flipped versions of each video clip were created using Adobe Premiere Pro, so that the speaker also seemed to be gesturing with her non-dominant hand. This was to balance out a potential right-hand bias.

Once the clips had been recorded, gestures were coded to control for several features to ensure that these were evenly distributed across the lists and conditions (see Appendix S1, Supplementary Materials). Semantically related gestures were coded and controlled for viewpoint (Character versus Observer Viewpoint; McNeill, 1992). A character viewpoint incorporates the speaker's body into gesture space, with the speaker's hands representing the hands of a character: e.g., the speaker might move her hand as if she were slicing meat herself (Fig. 1b). In contrast, an observer-viewpoint gesture excludes the speaker's body from gesture space, and hands play the part of the character as a whole: the speaker might move her hand from left to right with a swinging movement to depict a character swinging on a rope (Fig. 1c).

Gestures were further coded for their timing relative to speech to ascertain that the stroke coincided temporally with the spoken verb form. We further coded gestures for 'single' versus 'repeated

[5]According to the International Association of Conference Interpreters AIIC. (2019a), the 'A' language is the interpreter's mother tongue (or the language they speak best; it is an active language into which they work from all their other working languages). A 'B' language is also an active language, in which the interpreter is perfectly fluent, but is not their mother tongue. A 'C' language is a passive language which the interpreter understands perfectly but into which they do not work – they will interpret from this language into their active language(s).

[6]3 participants had two A languages: French and English

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

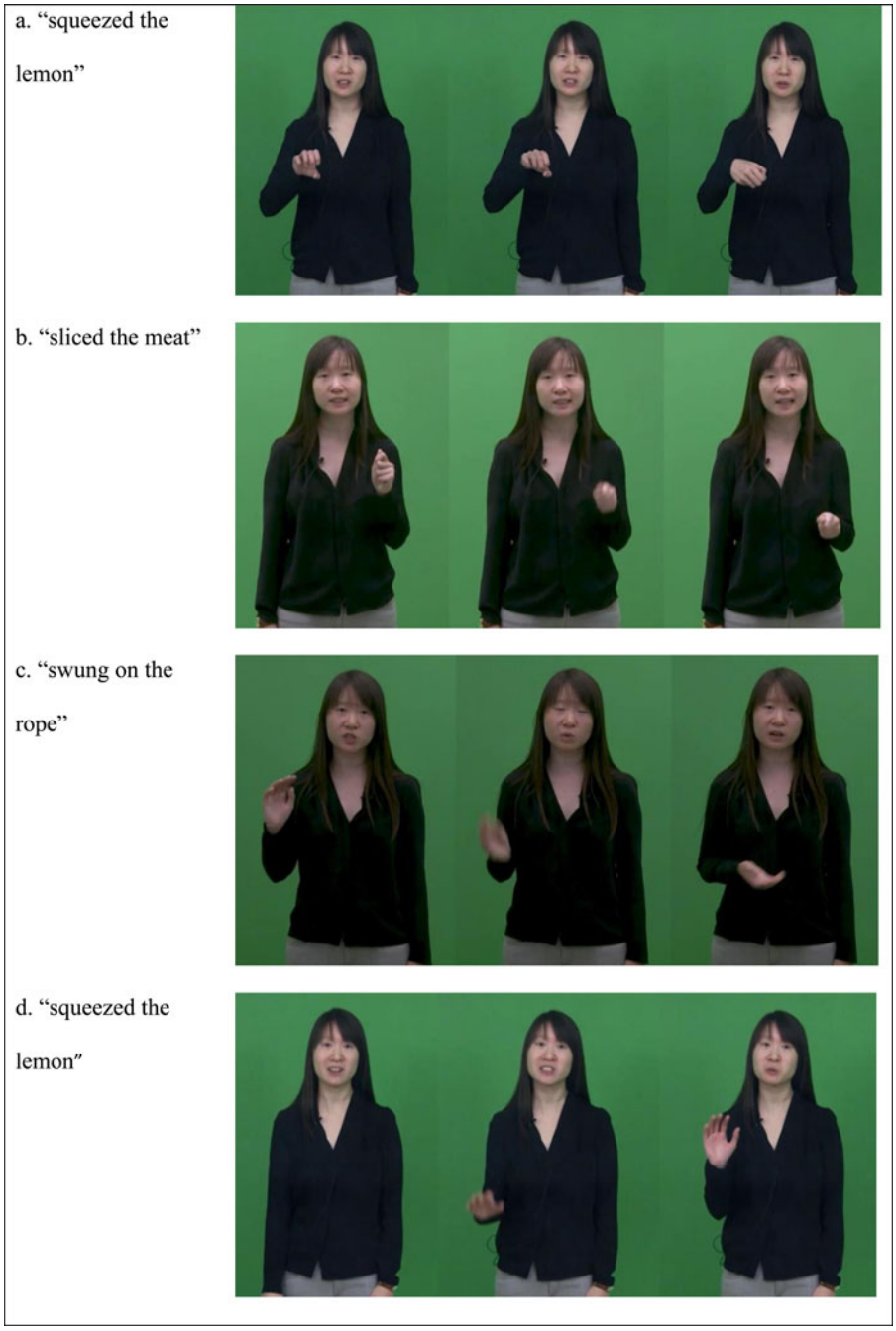


Fig. 1. Examples of gestures. a. Character-viewpoint semantically related gesture for "squeezing". b. Character-viewpoint semantically related gesture for "slicing". c. Observer-viewpoint semantically related gesture for "swinging". d. Semantically unrelated gesture for "squeezing".

stroke'. In single stroke gestures the stroke is performed once, while in repeated gestures the stroke is repeated twice. The number of repetitions was matched across lists for semantically related and unrelated gestures. Place of gestural articulation was coded following an adapted version of McNeill's schema of gesture space (McNeill, 1992, p. 89) as in Gullberg and Kita (2009). The 'center-center' and 'center' categories were merged into one 'center' category, while the 'upper periphery', 'lower periphery', etc., were merged into one 'periphery' category. Place of articulation was thus coded as either 'center', 'periphery' or 'center-periphery'. Gestures were also coded for complexity of trajectory. Straight lines in any direction were coded as a 'simple trajectory' and more complex patterns were coded as 'complex trajectories' (e.g., when the stroke included a change of direction).

Verb duration was determined for each video clip by identifying verb onset, offset and preposition offset in the case of the Observer-Viewpoint category. Mean verb duration was comparable between semantically related items ($M = 491$ ms, $SD = 98$) and semantically unrelated items ($M = 496$ ms, $SD = 112$). Mean verb duration of no-gesture items was significantly shorter ($M = 445$ ms, $SD = 112$) than both semantically related gesture items ($p < .05$) and semantically unrelated gesture items ($p < .05$), possibly because the coordination of speech and gesture slowed down production.

Stroke duration was determined for all gestures and included post-stroke-holds, when present. Mean stroke duration did not differ significantly between semantically related ($M = 585$ ms, $SD = 118$) and unrelated gestures ($M = 612$ ms, $SD = 152$).

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

Semantically related and unrelated items were comparable in terms of stroke type and of complexity of trajectory. Both categories included 67% single stroke gestures (40 items) versus 33% repeated stroke gestures (20 items), and 83% simple trajectories (50 gestures) versus 17% complex trajectories (10 gestures). Items differed in place of articulation, as semantically unrelated items were mostly articulated in the 'center-periphery' area (65%, 39 items) whereas semantically related gestures were mostly performed centrally (58%, 35 gestures).

All gestures used in the experiment are described in Appendix S2 (Supplementary Materials).

Pictures

Black-and-white line drawings corresponding to the actions depicted in the target verbs were taken from the IPNP database (Szekely, Jacobsen, D'Amico, Devescovi, Andonova, Herron, Lu, Pechmann, Pléh, Wicha, Federmeier, Gerdjikova, Gutierrez, Hung, Hsu, Iyer, Kohnert, Mehotcheva, Orozco-Figueroa, Tzeng, Tzeng, Arévalo, Vargha, Butler, Buffington, & Bates, 2004). Since more drawings were needed, most of the pictures were created by an artist using the same format. The drawings were normed for name and concept agreement, familiarity and visual complexity as in Snodgrass and Vanderwart (1980) by 11 L1 English speakers and 10 L1 French speakers. Pictures that did not yield satisfactory measures were redrawn and normed by 10 L1 English speakers and 11 L1 French speakers (some raters were involved in both norming rounds; total $n = 31$). A sweepstake incentive of 50 CHF (for each language group) was made available.

Raters were asked to identify pictures as briefly and unambiguously as possible by typing in the first description (a verb) that came to mind. Concept agreement, which takes into account synonyms (e.g., "cut" and "carve" are acceptable answers for the target "slice"), was calculated as in Snodgrass and Vanderwart (1980). Picture pairs with concept agreement of over 70% were used. The same raters judged the familiarity of each picture – that is, the extent to which they came in contact with or thought about the concept. Concept familiarity was rated on a 5-point Likert-type scale (from 1 = "very unfamiliar" to 5 = "very familiar"). The same raters rated the complexity of each picture – that is, the amount of detail or intricacy of the drawings. Picture visual complexity was rated on a 5-point Likert-type scale (from 1 = "very simple" to 5 = "very complex").

As shown in Appendix S1, Supplementary Materials, lists were balanced in terms of sentence plausibility, verb frequency, verb duration, gesture viewpoint, stroke type, stroke duration, place of articulation, gesture trajectory, concept agreement, concept familiarity, and visual complexity of the picture. Gesture conditions were balanced as to stroke type, stroke duration, gesture trajectory, but differed in terms of verb duration and place of articulation.

We created 24 blocks to accommodate the three gesture conditions, and counterbalance for gesture handedness (right/left hand), and target picture position (right/left side). Each block comprised four practice trials and 30 critical trials (10 of each gesture condition). Trial-type order was randomised in each block. Each session consisted of four blocks, two assigned to the SI activity, two to the passive viewing/listening activity. Activity order was counterbalanced across participants. Each participant saw List A and List B twice, but never saw the same individual trial twice. A total of 180 sentences were created, and each participant was presented with 60 experimental sentences in each task, meaning with 120 sentences in the whole experiment. Of these 120 sentences, one third corresponded to the condition without any gestures, one third to the semantically related gesture condition and one third to the semantically unrelated gesture condition.

Apparatus

Experimental tasks were completed in an ISO4043-compliant mobile interpreting booth (ISO, 2016a), programmed in SR Experiment-Builder® and deployed on a Mac Mini®. Visual stimuli were presented on a 23" (58.4 cm) HP E232 display with a refresh rate of 60 Hz, located approximately 75 cm from the participants. Auditory stimuli were played over an LBB 3443 Bosch headset. Eye-movement data were acquired with an SR Research EyeLink® 1000 desktop-mounted remote eye-tracking system with a sampling rate of 500 Hz. The eye-tracker camera was located in front of the monitor, leaving a distance of approximately 60 cm between participants' eyes and the eye-tracker. Participants' spoken interpretations were recorded using a Bosch DCN-IDESK-D interpreting console and fed back into the EyeLink to generate time-aligned stereophonic recordings of stimulus audio output and participant audio input. The input device was a VPixx Technologies RESPONSEPixx HANDHELD 5-button response box.

Procedure

Each session consisted of four blocks and lasted approximately one hour. Each block started with a standard 9-point calibration of the eye-tracker. After validation, participants completed a practice-trial session. During and at the end of the practice session, participants could ask questions. Participants were then instructed to launch the critical trials by pressing a button. Participants had timed three-minute breaks between blocks. During interpreted blocks, the experimenter monitored whether participants were interpreting the trial sentences simultaneously, and, if necessary, reminded participants. No feedback was given during the experiment. The experimenter monitored the eye-tracking display and recalibrated when necessary.

Passive viewing/listening activity – picture-matching task

Participants were asked to "keep looking at the screen while the video [was] being played" to enable the eye-tracker to follow their gaze. They were instructed to use the response box to "choose the picture that best correspond[ed] to the video" between two pictures. Upon launch of a trial, the participants saw a short video clip as described in the Task and materials section. This was followed by a blank screen (2,000 ms) upon which two pictures were presented, respectively on the left and right side of the screen. Once a picture was selected, a drift correction was performed to proceed to the next trial. The procedure is illustrated in Figure 2.

SI activity – picture-matching task

Participants were asked to "start interpreting as soon as possible when the video start[ed]", so that they would be engaged in simultaneous interpreting by the time the target verb was uttered. They were also instructed to "keep looking at the screen while the video [was] being played" to enable the eye-tracker to follow their gaze. They were asked to use the response box to "choose the picture that best correspond[ed] to the video" between two pictures. Upon launch of a trial, the participants saw a short video clip as described in the Task and materials section. This

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

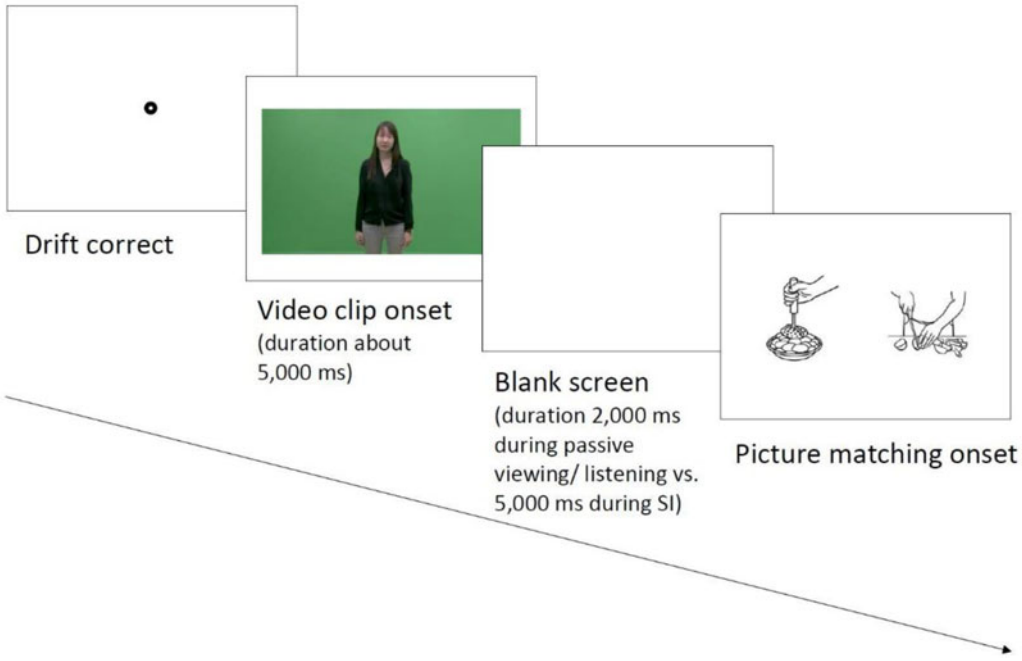


Fig. 2. Trial sequence during the passive viewing/listening versus SI activities.

was followed by a blank screen (5,000 ms, which gave participants time to complete their interpretation) upon which two pictures were presented, respectively on the left and right side of the screen. Once a picture was selected, a drift correction was performed to proceed to the next trial. The procedure is illustrated in Figure 2.

Analysis

The analyses for the three dependent variables, response accuracy, reaction time (RT) and dwell time, were conducted separately and implemented in R (R Core Team, 2013) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015).

Practice trials were not included in the analyses. Trials in which participants had not interpreted, only partially interpreted, or had not finished interpreting the stimuli by the onset of the picture-selection task were also excluded from the analysis, which led to the removal of 13.5% of interpreted trials (195 trials, 6.8% of the whole dataset). As one participant had systematically pressed the central button rather than the left or right button in the picture-matching task, the first two blocks of this testing session were excluded (instructions were followed after that).

Accuracy

Accuracy data were analysed using generalised linear mixed models (GLMM). The dataset was trimmed before completing the analyses. Responses above and below 3 SDs from the RT mean were considered outliers, which led to the removal of 2.7% (34 trials) of the data points in the SI activity dataset, and 2% (28 trials) of the passive viewing/listening activity dataset. Overall, 10% (287 trials) of all trials were excluded in the Accuracy data analyses[7].

GLMM analyses were conducted to test the relationship between accuracy and the fixed effects *activity* (2 levels, passive viewing/listening and simultaneous interpreting) and *semantic match between speech and gesture* (3 levels, semantically related gesture, semantically unrelated gesture, no gesture). An interaction term was set between activity type and semantic match. Subjects and items were entered as random effects with by-subject and by-item random intercepts as this was the maximal random structure supported by the data.

Reaction time

Linear mixed-effects model (LMM) analyses were run on RT data. Significance of effects was determined by assessing whether the associated *t*-statistics had absolute values ≥ 2. The dataset was trimmed before completing the analyses using the same approach as for the Accuracy data. Only accurate trials were used for the RT analyses, resulting in the exclusion of another 2.3% (60 trials) of RT data points. Overall, the excluded trials amounted to 12% (347 trials) of all trials[8].

RTs were log-transformed and analysed using a LMM with the same fixed-effects structure as the GLMM. Subjects and items were entered as random effects with by-subject and by-item random intercepts.

Dwell time

LMM analyses were run on dwell time data. The same values as for RT data were used to determine significance of effects. Dwell time analyses were only performed on the two conditions that contained any gestures (66.6% of the data). Only accurate trials were used, resulting in the exclusion of 2.9% (52 trials) of the total data points. Two areas of interest were created, one comprising the speaker's head, the other one including gesture space, from the speaker's shoulders to her hips. Dwell time (in ms) was measured in each of the areas of interest during the gesture stroke. We tested the relationship between accuracy and the fixed effects

[7]The difference between the cumulated excluded trials and the grand total stems from the blocks excluded as one participant did not follow instructions

[8]See above

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

Table 2. (A) Mean response-accuracy percentages. (B) Mean RT in ms.

	Passive viewing/ listening cond.		Simultaneous interpreting cond.	
Accuracy	*M*	*SD*	*M*	*SD*
Semantically related gesture	97.7	2.6	97.9	3.2
Semantically unrelated gesture	97.7	3.3	97.1	3.9
No gesture	97.5	3.3	98.4	3.2
RT	Passive viewing/ listening cond.		Simultaneous interpreting cond.	
	M	*SD*	*M*	*SD*
Semantically related gesture	1,405	581	1,437	632
Semantically unrelated gesture	1,529	716	1,432	562
No gesture	1,491	704	1,451	665

activity (2 levels, passive viewing/listening and SI) and *semantic match between speech and gesture* (2 levels, semantically related gesture versus semantically unrelated gesture). An interaction term was set between activity type and semantic match. Subjects and items were entered as random effects with by-subject and by-item random intercepts as this was the maximal random structure supported by the data.

Results

Accuracy

Accuracy scores (Table 2A) were close to ceiling in both activities and all conditions.

The no-gesture condition and the passive viewing/listening activity were set as baselines. The interaction between activity type and gesture condition was not significant ($\beta = -0.57$, $SE = 0.75$, $Z = -0.76$, $p = .45$), indicating that the two fixed effects did not interact to affect accuracy. The result of the likelihood-ratio test used to compare the full to the reduced model (without interaction term) was also not significant (χ^2 (2) = 2.83, $p = .24$), confirming this result. The reduced model, which thus provided a better fit to the data, revealed that accuracy was not affected by activity type or gesture condition individually (SI: $\beta = 0.03$, $SE = 0.28$, $Z = 0.10$, $p = .92$; semantically unrelated gesture: $\beta = -0.34$, $SE = 0.34$, $Z = -0.99$, $p = .32$; semantically related gesture: $\beta = -0.11$, $SE = 0.35$, $Z = -0.32$, $p = .75$).

Setting the semantically unrelated gesture condition as baseline to explore potential effects of semantically related as compared to semantically unrelated gestures, using the relevel function in R, the interaction between activity type and gesture condition was not significant either ($\beta = 0.61$, $SE = 0.66$, $Z = 0.93$, $p = .35$), indicating that the two fixed effects did not interact to affect accuracy. The reduced model revealed that accuracy was also not affected by gesture condition individually (no-gesture: $\beta = 0.34$, $SE = 0.34$, $Z = 0.99$, $p = .32$; semantically related gesture: $\beta = 0.23$, $SE = 0.33$, $Z = 0.69$, $p = .49$).

Reaction time

Mean reaction times are presented in Table 2B.

The no-gesture condition and the passive viewing/listening activity were set as baselines. The interaction between activity type and gesture condition was not significant ($\beta = 0.03$, $SE = 0.04$, $t = 0.92$), indicating that the two variables did not interact to affect RTs. The result of the likelihood-ratio test used to compare the full to the reduced model (without interaction term) was also not significant (χ^2 (2) = 1.76, $p = .41$), confirming this result. The output of the reduced model revealed that RTs were not affected by activity type ($\beta = -0.01$, $SE = 0.01$, $t = -0.34$) or gesture condition (semantically unrelated gesture: $\beta = 0.01$, $SE = 0.02$, $t = 0.82$; semantically related gesture: $\beta = -0.03$, $SE = 0.02$, $t = -1.52$).

Setting the semantically unrelated gesture condition as baseline to explore potential effects of semantically related as compared to semantically unrelated gestures, the interaction between activity type and gesture condition was not significant either ($\beta = 0.05$, $SE = 0.04$, $t = 1.29$), indicating that the two fixed effects did not interact to affect RT. However, the reduced model revealed that RT was affected by gesture condition individually with semantically related gestures significantly affecting RTs (no gesture: $\beta = -0.01$, $SE = 0.02$, $t = -0.82$; semantically related gesture: $\beta = -0.04$, $SE = 0.02$, $t = -2.34$).

Dwell time

Visual attention to gesture was low (see Figure 3). This is in line with the literature: the speaker's face dominates as a kind of "default location" and addressees look directly at very few gestures (Gullberg & Holmqvist, 2006; Gullberg & Kita, 2009). This is also in line with what we know of interpreters' preference for the speaker's face during SI (Seeber, 2011).

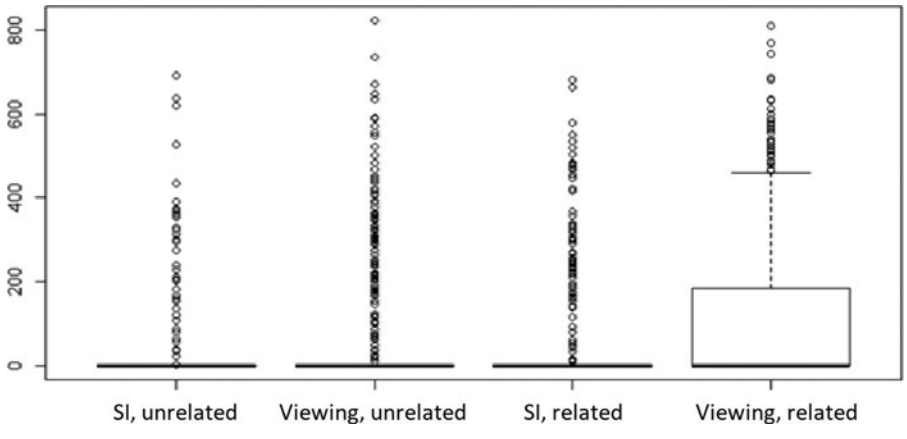


Fig. 3. Dwell time on gesture space in ms according to gesture condition and activity type (SI = simultaneous interpreting, viewing = passive viewing/listening, unrelated = semantically unrelated gesture, related = semantically related gesture).

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

The passive viewing/listening activity and the control gesture condition were set as baselines. The interaction between activity type and gesture condition was significant ($\beta = -27.13$, $SE = 12.09$, $t = -2.24$), indicating that activity type and gesture condition interacted to affect dwell time on the speaker's gesture space. The result of the likelihood-ratio test used to compare the full to the reduced model was also significant (χ^2 (1) = 5.03, $p = .03$), indicating that the full model was a better fit to the data than the reduced model. The model output indicated that dwell time was significantly affected both by activity type (SI: $\beta = -46.43$, $SE = 6.13$, $t = -7.58$) and gesture condition (semantically related gesture: $\beta = 32.98$, $SE = 6.14$, $t = 5.37$).

Discussion

Accuracy was not affected by activity type, gesture condition or any interaction thereof. Thus, it appears that neither semantically related gestures nor semantically unrelated gestures had an effect on interpreters' accuracy in either activity. That said, 13.5% of the interpreted trials had to be excluded from analysis since interpretations either had not been completed, were incomplete, or had not been completed by the time pictures were displayed. Stimuli that took interpreters more time to interpret or generated incomplete interpretations may have been associated with more difficulty. This might have caused higher error rates in the picture-matching task. Since the audio recordings stopped when the pictures were presented, however, a post-hoc analysis of these trials is impossible.

Activity type did not affect RTs. Nor was there any evidence that semantically related speech-gesture pairs made interpreters faster (in either activity) compared to utterances without gestures. When interpreters were presented with either semantically related or semantically unrelated gestures, however, semantically related gestures were associated with faster RTs (in either activity) than semantically unrelated gestures. This RT difference suggests that interpreters integrated gestures and that language comprehension was sensitive to gestures' semantic relationship to the spoken utterance. It also raises the question of whether semantically related speech-gesture pairs accelerated comprehension or whether semantically unrelated speech-gesture pairs slowed it down compared to the baseline. Collapsed across activities, mean RTs were fastest in the semantically related gesture condition ($M = 1{,}420$ ms, $SD = 605$), slower in the no-gesture condition ($M = 1{,}472$ ms, $SD = 686$), and slightly slower still in the semantically unrelated gesture condition ($M = 1{,}484$ ms, $SD = 651$). Mean RTs in the no-gesture condition and the semantically unrelated gesture condition were very similar, and the difference between the semantically related gesture condition and the no-gesture condition approached significance in the LMM. This rather points to an acceleration effect of semantically related gestures than to a slow-down effect of semantically unrelated gestures compared to the baseline.

The dwell time measure points in the same direction. Interpreters attended to semantically related gestures significantly longer than to semantically unrelated gestures in both activities, which suggests that interpreters' visual attention patterns, too, are sensitive to the semantic relationship between gesture and speech. Therefore, the gestures did not simply attract participants' attention irrespective of their relevance in the utterance (as in Rayner, 1998).

Interpreters attended to gestures significantly longer during the passive viewing/listening activity than during SI, and dwell time was highest when interpreters attended to semantically related gestures during passive viewing/listening. This may reflect task demands in SI. However, they did attend longer to semantically related than to semantically unrelated gestures in this activity, too, which suggests that interpreters' preference for the speaker's face during SI did not prevent them from attending to and integrating gestures, taking into account their semantic relationship with the utterance.

The experiment did not bring to light any language comprehension differences between passive viewing/listening and SI in terms of accuracy and reaction time: therefore, engaging in SI did not modulate language comprehension. However, interpreters might have honed their cognitive abilities due to their experience of SI, and interpreting experience may have had an effect on the interpreters' behaviour. Other bilinguals without interpreting experience might behave differently from the tested interpreters, since interpreting expertise has been shown to positively influence cognitive performance, e.g., dual-task performance (Strobach et al., 2015), and cognitive flexibility (Yudes et al., 2011). To investigate this, a second experiment compared interpreters to bilinguals without interpreting experience.

Second experiment

Method

Participants

The second experiment examined passive viewing/listening only in an experimental group consisting of the interpreters in the first experiment and a comparison group of professional translators without interpreting experience. We compared simultaneous interpreters with translators since the two groups are likely to be similar in terms of language proficiency and age and are used to working with two languages. The groups were matched for factors pertaining to background and language experience (see Table 3).

Twenty-four translators working from English into French participated in the experiment[9]. They were recruited via an e-mail describing the eligibility criteria, and interested individuals were invited to sign up for the experiment. They completed a questionnaire similar to the one used in the first experiment, with adapted questions regarding their professional background. The group included two participants who were trained translators no longer working in this field but in related fields (e.g., lecturer). The remaining 22 participants had been pursuing a career in translation for a mean of 10 years; 7 months ($SD = 9$ years; 10 months). Four participants had previously received training in conference interpreting. However, they had never worked as professional interpreters. All participants had normal or corrected-to-normal vision and reported no language disorders. Their L1 was French[10], their L2 English[11] (which could be a passive or an active language, similarly to the interpreters' group).

The translators were younger and less experienced than the interpreters. Interpreters also rated their listening ability in English as higher than translators. It is likely that the difference in perceived listening proficiency is linked to the different professional profiles of the two groups, since L2 listening skills are a key aspect of SI.

[9]One additional participant could not be tested as calibration could not be performed.

[10]$n = 22$ as this question related to the translator's language combination. For all measures on language abilities, $n = 24$ as these were general questions not regarding participants' professional background.

[11]Same as above

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

Table 3. Background information provided in the language background questionnaire, and comparison of groups (t-test for numerical variables, Wilcoxon test for ordinal variables): *: p < 0.05, **: p < 0.01, *** : p < .001.

	Interpreters (*n* = 24)		Translators (*n* = 24)		
	M	*SD*	*M*	*SD*	*Comparison*
Background					
Age *(yrs)*	47.38	11.37	38.46	10.26	**
Languages spoken	4.75	1.45	3.96	0.86	*
Professional experience *(yrs)*	19.25	11.00	10.59 *(n = 22)*	9.84	**
French language					
Age of Acquisition *(yrs)*	2.65	3.22	0.75	0.98	**
Age of Fluency *(yrs)*	4.58	3.39	3.5	1.85	
Duration of exposure[12] *(yrs)*	39.87	12.42	36.23	10.08	
Current exposure *(%)*	54.95 *(n = 22)*	16.07	66.95 *(n = 22)*	15.64	*
English language					
Age of Acquisition *(yrs)*	8.77	4.44	10.75	2.95	
Age of Fluency *(yrs)*	17.42	6.83	19.88	4.64	
Duration of exposure[13] *(yrs)*	2.46	3.42	3.23	8.68	
Current exposure *(%)*	20.86 *(n = 22)*	8.23	18.27 *(n = 22)*	10.54	
Self-rated English proficiency					
Speaking	7.83	1.49	7.83	1.01	
Reading	9.17	0.70	8.42	0.78	**
Listening	8.88	0.74	7.92	1.21	**

All participants gave written informed consent. The experiment was approved by the Faculty of Translation and Interpreting's Ethics Committee. No participant was involved in the norming of the stimuli.

Design, task, materials, procedure

Participants completed the passive viewing/listening activity only. The task, materials, apparatus, procedure and instructions were the same as in the first experiment. Participants completed four blocks, following the same rotation as in the first experiment. However, since there was only one activity, the analysis included only two of the blocks, those corresponding to the passive viewing/listening blocks in the first experiment. The sessions lasted approximately 50 minutes. Thus, time on task for translators was slightly less than the interpreters (viewing/listening trials were slightly shorter as no margin had to be added for interpretations). They saw each list twice, like the interpreters, even though they completed the same activity four times whereas the interpreters completed two different activities twice.

Analysis

The dependent variables were the same as in the first experiment, and analyses were conducted separately.

Accuracy

The dataset was trimmed before completing the analyses. Responses above and below 3 SDs from the RT mean were considered outliers, which led to the removal of 1.9% (27 trials) of the interpreters' dataset, and 2.8% (40 trials) of the translators' dataset. GLMM analyses were conducted to test the relationship between accuracy and the fixed effects *group* (2 levels, interpreter or translator status) and *semantic match between speech and gesture* (3 levels, semantically related gesture, semantically unrelated gesture, no gesture). An interaction term was set between group and semantic match. Subjects and items were entered as random effects with by-subject and by-item random intercepts as this was the maximal random structure supported by the data.

Reaction time

The dataset was trimmed before completing the analyses using the same approach as for the Accuracy data. Only accurate trials were analysed, resulting in the exclusion of another 2.6% (72 trials) of the data points. Overall, the excluded trials amounted to 5.9% (169 trials) of the total[14].

RTs were log-transformed, and analysed using a LMM with the same fixed-effects structure as the GLMM. Subjects and items were entered as random effects with by-subject and by-item random intercepts.

Dwell time

Dwell time analyses were only performed on the two conditions that contained any gestures (66.6% of the data). Only accurate

[12]Number of years participants spent in a country or region where the relevant language is spoken

[13]See above

[14]The difference between the cumulated excluded trials and the grand total stems from the blocks excluded as one participant did not follow instructions

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

Table 4. (A) Mean response-accuracy percentages. (B) Mean RT in ms.

	Interpreters		Translators	
Accuracy	*M*	*SD*	*M*	*SD*
Semantically related gesture	97.7	2.6	98.5	5.2
Semantically unrelated gesture	97.7	3.3	97.2	3.7
No gesture	97.5	3.3	96.0	5.1
	Interpreters		Translators	
RT	*M*	*SD*	*M*	*SD*
Semantically related gesture	1,405	581	1,506	730
Semantically unrelated gesture	1,529	716	1,580	811
No gesture	1,491	704	1,515	750

trials were used, resulting in the exclusion of 3.2% of the dataset. The same areas of interest as in the first experiment were used. Dwell time data were analysed using a LMM to test the relationship between dwell time and the fixed effects *group* (2 levels, interpreter or translator status) and *semantic match between speech and gesture* (2 levels, semantically related gesture or semantically unrelated gesture). An interaction term was set between group and semantic match. Subjects and items were entered as random effects with by-subject and by-item random intercepts.

Results

Accuracy

Accuracy scores (Table 4A) were close to ceiling in both groups and in all conditions.

The no-gesture condition and the translator group were set as baselines. The interaction between group membership and gesture condition was not significant ($\beta = -1.20$, $SE = 0.68$, $Z = -1.76$, $p = .08$), indicating that the two fixed effects did not interact to affect accuracy. The result of the likelihood-ratio test used to compare the full to the reduced model (without interaction term) was also not significant (χ^2 (2) = 3.36, $p = .19$), confirming this result. The reduced model, which thus provided a better fit to the data, revealed that whereas accuracy was not affected by group membership, gesture condition had a significant effect on accuracy, with semantically related gestures significantly affecting accuracy (group membership: $\beta = 0.03$, $SE = 0.38$, $Z = 0.09$, $p = .93$; semantically unrelated gesture: $\beta = 0.32$, $SE = 0.31$, $Z = 1.03$, $p = .30$; semantically related gesture: $\beta = 0.70$, $SE = 0.33$, $Z = 2.09$, $p = .04$).

Setting the semantically unrelated gesture condition as baseline to explore potential effects of semantically related gestures as compared to semantically unrelated gestures, the interaction between group membership and gesture condition was not significant ($\beta = -0.97$, $SE = 0.71$, $Z = -1.37$, $p = .17$, indicating that the two fixed effects did not interact to affect accuracy. The reduced model revealed that accuracy was also not affected by gesture condition individually (no gesture: $\beta = -0.32$, $SE = 0.31$, $Z = -1.03$, $p = .30$; semantically related gesture: $\beta = 0.38$, $SE = 0.35$, $Z = 1.08$, $p = .28$).

Reaction time

Mean reaction times are presented in Table 4B.

The no-gesture condition and the translator group were set as baselines. The interaction between group membership and gesture condition was not significant ($\beta = -0.04$, $SE = 0.03$, $t = -1.12$), indicating that the two variables did not interact to affect RTs. The result of the likelihood-ratio test used to compare the full to the reduced model (without interaction term) was also not significant (χ^2 (2) = 1.26 ms, $p = .53$), confirming this result. The output of the reduced model revealed that RTs were not affected by group membership ($\beta = -0.03$, $SE = 0.07$, $t = -0.39$) or gesture condition (semantically unrelated gesture: $\beta = 0.03$, $SE = 0.02$, $t = 1.80$; semantically related gesture: $\beta = -0.02$, $SE = 0.02$, $t = -1.48$).

Setting the semantically unrelated gesture condition as baseline to explore potential effects of semantically related gestures as compared to semantically unrelated gestures revealed that the interaction between group membership and gesture condition was not significant ($\beta = -0.02$, $SE = 0.03$, $t = -0.64$, indicating that the two fixed effects did not interact to affect RT. However, the reduced model revealed that RTs were affected by gesture condition individually, with a significant effect of semantically related gestures (no gesture: $\beta = -0.03$, $SE = 0.02$, $t = -1.80$; semantically related gesture: $\beta = -0.05$, $SE = 0.02$, $t = -3.29$).

Dwell time

Visual attention to gesture was low, and semantically related gestures were fixated for longer than semantically unrelated gestures (see Figure 4).

The translator group and the control gesture condition were set as baselines. The interaction between group membership and gesture condition was not significant ($\beta = 11.02$, $SE = 13.27$, $t = 0.83$), indicating that the two variables did not interact to affect

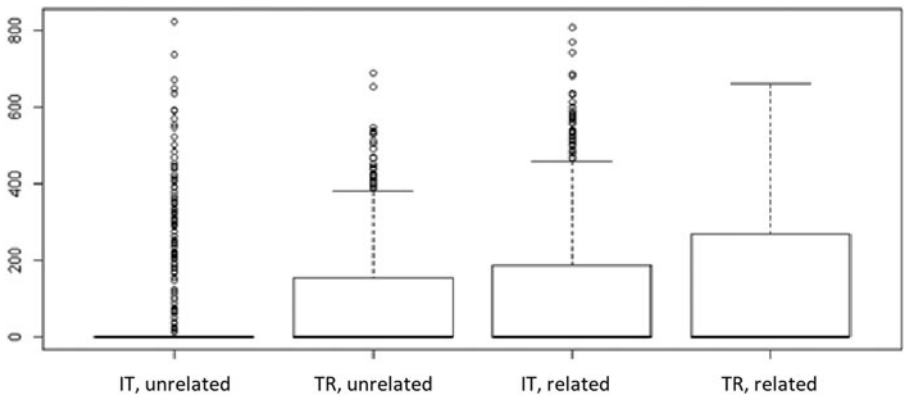


Fig. 4. Dwell time on gesture space in ms according to gesture condition and group membership (IT = interpreters, TR = translators, unrelated = semantically unrelated gesture, related = semantically related gesture).

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

dwell time. The result of the likelihood-ratio test used to compare the full to the reduced model was also not significant (χ^2 (1) = 0.69, p = .41), confirming this result. The output of the reduced model revealed that whereas dwell time was not affected by group membership (β = −23.86, SE = 24.49, t = −0.97), gesture condition had a significant effect on dwell time (semantically related gesture: β = 40.35, SE = 6.64, t = 6.08).

Discussion

The second experiment did not show any significant differences in language comprehension or overt visual attention between interpreters and translators, which suggests that interpreting experience did not affect interpreters' behaviour.

Accuracy was affected by gesture condition since both groups were significantly more accurate when presented with semantically related gestures than with audiovisual utterances without gesture. Semantically unrelated gestures did not have the same effect. Therefore, the gestures did not simply attract participants' attention irrespective of their relevance in the utterance (as in Rayner, 1998). Instead, the semantic relevance of gesture in the utterance seemed to have a clear effect.

No significant RT differences were found between the no-gesture and the semantically related gesture conditions. However, when participants were presented with semantically related or unrelated gestures, semantically related gestures were associated with faster RTs (in both groups) compared to semantically unrelated gestures. This suggests that both in interpreters and in translators, gestures are integrated and language comprehension during passive viewing/listening is sensitive to gestures' semantic relationship with the spoken utterance. As before, this raises the question of whether semantically related gestures accelerated comprehension or whether semantically unrelated gestures slowed it down compared to the baseline. However, the data do not allow us to determine this: collapsed across activities, mean RTs were fastest in the semantically related gesture condition (M = 1,456 ms, SD = 662), slower in the no-gesture condition (M = 1,503 ms, SD = 727), and slower still in the semantically unrelated gesture condition (M = 1,554 ms, SD = 765).

The dwell time measure points in the same direction as the RTs. Although few of the gestures were fixated, participants attended to the speaker's gesture space. Both groups attended to semantically related gestures significantly longer than to semantically unrelated gestures, which suggests that participants' visual attention patterns, too, are sensitive to the semantic relationship between gesture and speech.

General discussion

This study aimed to probe the potential effect of co-speech gestures on language comprehension in simultaneous interpreters. The first question was whether simultaneous interpreters integrate gestural information during language comprehension (RQ 1). Both during passive viewing/listening and SI, interpreters' language comprehension was faster with semantically related gestures compared to semantically unrelated gestures, and this was most likely attributable to an acceleration effect of semantically related gestures compared to the no-gesture baseline than to a slow-down effect of semantically unrelated gestures. Thus, language comprehension was sensitive to the semantic relevance of gestures in the utterance. We draw two conclusions from this. First, co-speech gestures are indeed integrated; they are part and parcel of language comprehension also in the extreme form of language use that is SI. Second, semantically related co-speech gestures have a facilitatory effect on simultaneous interpreters' language comprehension, just as in L1 and L2 language use (Dahl & Ludvigsen, 2014; Hostetter, 2011; Sueyoshi & Hardison, 2005). This is in contrast to statements in the interpreting literature, where most studies have found no significant effect of visual input on SI (Anderson, 1994; Bacigalupe, 1999; Balzani, 1990; Rennert, 2008; Tommola & Lindholm, 1995), perhaps because these studies presented a variety of visual cues to participants and/or compared audiovisual to audio-only conditions whereas, in the current study, the main variable was co-speech gestures and all conditions were audiovisual.

The second question was whether integration of gestural information during language comprehension is affected by task (during SI versus during passive viewing/listening) or interpreting experience (interpreters versus translators); RQs 2 and 4. The first experiment revealed no significant task effect on interpreters' behaviour across passive viewing and SI. SI is considered mentally taxing (Seeber, 2015) since it combines concurrent spoken language comprehension and production in two distinct languages. However, it seems that the language comprehension component in SI does share common features with other language comprehension tasks (Seeber, 2017), and that the fact that simultaneous interpreters produce a verbal response while comprehending the speaker's input does not modulate the effect of co-speech gestures on comprehension. The second experiment probed whether interpreting experience affected behaviour, comparing interpreters to bilinguals with no interpreting experience. Both groups' behaviour was similarly affected by gestures – they were significantly more accurate with semantically related gestures than with utterances without gestures, and were NOT more accurate when presented with semantically unrelated gestures compared to utterances without gestures. Moreover, language comprehension was significantly faster with semantically related gestures compared to semantically unrelated gestures, although we were not able to attribute this to an acceleration effect of semantically related gestures or to a slow-down effect of semantically unrelated gestures. Again, the results suggest that semantically related gestures improved language comprehension, in line with the literature (e.g., Dahl & Ludvigsen, 2014; Hostetter, 2011; Sueyoshi & Hardison, 2005).

Importantly, the experiment did not reveal any significant differences between groups, suggesting that interpreters and other bilinguals do not differ in how gestures impact comprehension. This is in contrast to other findings, according to which interpreting expertise might positively influence cognitive performance, e.g., dual-task performance (Strobach et al., 2015) or cognitive flexibility (Yudes et al., 2011), suggesting that interpreters might be better than other bilinguals at integrating visual and auditory input in parallel. However, no studies to date have systematically investigated the effect of co-speech gestures. In conclusion, neither the SI task nor interpreting experience modulate the effect of co-speech gestures on bilinguals' language comprehension.

The last question was whether simultaneous interpreters and bilinguals without SI experience visually attend to gestures and whether they visually attend to them in the same way (RQs 3 and 5). In the first experiment, interpreters attended to the speaker's gesture space both during passive viewing/listening and during SI, confirming that interpreters do overtly attend to visual input, including to co-speech gestures, as in Seeber (2011). Visual attention was modulated by the semantic speech-gesture

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

relationship, and visual information was integrated, as in Stachowiak-Szymczak (2019). However, most of interpreters' overt visual attention was focused on the speaker's face rather than on her gesture space during SI, as in previous eye-tracking studies of visual attention to gestures (Gullberg & Holmqvist, 2006; Gullberg & Kita, 2009), including in a SI context (Seeber, 2011). But interpreters looked significantly longer at the speaker's gesture space during passive viewing/listening than during SI. It may be that interpreters engaged in SI preferred fixating the speaker's face to glean verbal speech information (see Jesse, Vrignaud, Cohen, & Massaro, 2000). However, our areas of interest do not allow us to assess which facial cues participants might have attended to. The second experiment showed that bilinguals with no interpreting experience overtly attend to co-speech gestures similarly to interpreters during passive viewing/listening. Although few gestures were directly fixated, participants' overt visual attention patterns were equally sensitive to the semantic relationship between gesture and speech. Thus, overt visual attention to co-speech gestures is modulated by gestural characteristics, which is line with the literature (Beattie et al., 2010; Gullberg & Holmqvist, 1999, 2006; Gullberg & Kita, 2009).

A few final remarks are in order. The study only tested participants working from English to French. Further research needs to establish whether our findings can be generalised to other languages and varying experience working with input of varying quality. Moreover, recruitment constraints did not allow us to fully match interpreters and translators, notably in terms of age, professional experience and self-rated English Listening proficiency. Age, notably, could have had an effect on the studied variables. Even if older, more experienced participants were to be found, the difference in listening proficiency is most likely due to different professional demands and it may not be possible to fully match groups on this measure. Moreover, the sample size in the current study is limited, which calls for similar studies testing more participants. That said, this sample size is in line with the interpreting studies literature, and given that the International Association of Conference Interpreters counts about three thousand members worldwide, all languages considered (AIIC, 2019b), it still enables us to draw conclusions about the studied population.

Conclusion

We conclude that simultaneous interpreters' language comprehension and overt visual attention are sensitive to speakers' co-speech gestures and their semantic relationship with the utterance. Further, co-speech gestures can have a facilitatory effect on interpreters' language comprehension, and this effect is modulated neither by the SI task (e.g., by the fact that interpreters produce a verbal output whilst engaging in language comprehension) nor by interpreting experience. Taken together, this suggests that co-speech gestures are part and parcel of language comprehension in bilingual processing even in 'extreme bilingual language use', such as SI. It also demonstrates that the language comprehension component in SI shares common features with other language comprehension tasks. Overall, the results strengthen the case for SI to be considered a multimodal phenomenon (Galvão & Rodrigues, 2010; Seeber, 2017; Stachowiak-Szymczak, 2019) and to be studied, taught and practiced as such.

Supplementary Material. Supplementary material can be found online at https://doi.org/10.1017/S136672892200058X

S1. Characteristics of the stimuli. Description: Table 1 describes and compares lists in terms of sentence, gesture and picture criteria. Table 2 describes and compares gesture conditions in terms of sentence and gesture criteria. (Word, 15 Ko)

S2. Gesture description. Description: The spreadsheet describes semantically related and unrelated gestures, both for critical and practice trials. (Excel, 21.5 Ko)

Competing interests. The authors declare none.

Data availability. The data that support the findings of this study are available from the authors.

References

AIIC. (2007) *Checklist for simultaneous interpretation*. Retrieved from https://aiic.org/document/4395/Checklist%20for%20simultaneous%20interpretation%20equipment%20-%20ENG.pdf

AIIC. (2019a) *The AIIC A-B-C*. Retrieved from https://aiic.org/site/world/about/profession/abc

AIIC. (2019b) *Inside AIIC*. Retrieved from https://aiic.org/site/world/about/inside

Anderson L (1994) Simultaneous interpretation: Contextual and translation aspects. In Lambert S and Moser-Mercer B (eds), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, pp. 101–120.

Bacigalupe LA (1999) Visual contact in simultaneous interpretation: Results of an experimental study. In Alvarez Lugris A and Fernandez Ocampo A (eds), *Anovar/anosar. Estudios de traduccion e interpretacion*. Vigo: Universidade de Vigo, Vol. 1, pp. 123–137.

Balzani M (1990) Le contact visuel en interprétation simultanée: résultats d'une expérience (Français–Italien). In Gran L and Taylor C (eds), *Aspects of applied and experimental research on conference interpretation*. Udine: Campanotto, pp. 93–100.

Bates D, Mächler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4 [sparse matrix methods; linear mixed models; penalized least squares; Cholesky decomposition]. *Journal of Statistical Software* **67**, 1–48. https://doi.org/10.18637/jss.v067.i01

Beattie G, Webster K and Ross J (2010) The fixation and processing of the iconic gestures that accompany talk. *Journal of Language and Social Psychology*, **29**, 194–213. https://doi.org/10.1177/0261927×09359589

Bühler H (1985) Conference interpreting: A multichannel communication phenomenon. *Meta*, **30**, 49–54.

Campana E, Silverman L, Tanenhaus MK, Bennetto L and Packard S (2005) *Real-time integration of gesture and speech during reference resolution*. Paper presented at the 27th annual meeting of the Cognitive Science Society.

Capirci O and Volterra V (2008) Gesture and speech: The emergence and development of a strong and changing partnership. *Gesture*, **8**, 22–44. https://doi.org/10.1075/gest.8.1.04cap

Cassell J, McNeill D and McCullough K.-E (1999) Speech–gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition*, **7**, 1–34. https://doi.org/10.1075/pc.7.1.03cas

Church RB, Garber P and Rogalski K (2007) The role of gesture in memory and social communication. *Gesture*, **7**, 137–158. https://doi.org/10.1075/gest.7.2.02bre

Colletta J-M, Guidetti M, Capirci O, Cristilli C, Demir OE, Kunene-Nicolas RN and Levine S (2015) Effects of age and language on co-speech gesture production: An investigation of French, American, and Italian children's narratives. *Journal of Child Language*, **42**, 122–145. https://doi.org/10.1017/S0305000913000585

Dahl TI and Ludvigsen S (2014) How I see what you're saying: The role of gestures in native and foreign language listening comprehension. *The Modern Language Journal*, **98**, 813–833. https://doi.org/https://doi.org/10.1111/modl.12124

Davies M (2008) The Corpus of Contemporary American English (COCA): One billion words, 1990-2019. Available online at http://corpus.byu.edu/coca/

Debras C (2017) The Shrug: Forms and Meanings of a Compound Enactment. *Gesture*, **16**, 1–34. https://doi.org/10.1075/gest.16.1.01deb

Frauenfelder U and Schriefers H (1997) A psycholinguistic perspective on simultaneous interpretation. *Interpreting*, **2**, 55–89. https://doi.org/https://doi.org/10.1075/intp.2.1-2.03fra

Galvão EZ (2013) Hand gestures and speech production in the booth: Do simultaneous interpreters imitate the speaker? In Carapinha C and Santos IA (eds), *Estudos de linguística*. Coimbra: Imprensa da Universidade de Coimbra, Vol. II, pp. 115–130. https://digitalis.uc.pt/handle/10316.2/32258

Galvão EZ and Rodrigues IG (2010) The importance of listening with one's eyes: A case study of multimodality in simultaneous interpreting. In Diaz Cintas J, Matamala A and Neves J (eds), *New insights into audiovisual translation and media accessibility*. Boston, USA: Brill | Rodopi, pp. 239–253. https://doi.org/https://doi.org/10.1163/9789042031814_018

Gieshoff A.-C (2018) *The impact of audio-visual speech input on work-load in simultaneous interpreting*. (Doctoral dissertation), Johannes Gutenberg-Universität Mainz, Mainz, Germany. Retrieved from https://publications.ub.uni-mainz.de/theses/volltexte/2018/100002183/pdf/100002183.pdf

Goldin-Meadow S (2003) *Hearing gesture: How our hands help us think*. Cambridge, MA, US: Belknap Press of Harvard University Press.

González LP (2014) Multimodality in translation and interpreting studies: Theoretical and methodological perspectives. In Bermann S and Porter C (eds), *A companion to Translation Studies*. Chichester: Wiley-Blackwell, pp. 119–131. https://doi.org/https://doi.org/10.1002/9781118613504.ch9

Graziano M and Gullberg M (2018) When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in psychology*, **9**. https://doi.org/10.3389/fpsyg.2018.00879

Gullberg M (2006) Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *International Review of Applied Linguistics in Language Teaching*, **44**, 103–124. https://doi.org/doi:10.1515/IRAL.2006.004

Gullberg M and Holmqvist K (1999) Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics & Cognition*, **7**, 35–63. https://doi.org/https://doi.org/10.1075/pc.7.1.04gul

Gullberg M and Holmqvist K (2006) What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, **14**, 53–82. https://doi.org/10.1075/pc.14.1.05gul

Gullberg M and Kita S (2009) Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of nonverbal behavior*, **33**, 251–277. https://doi.org/10.1007/s10919-009-0073-2

Hervais-Adelman A, Moser-Mercer B and Golestani N (2015) Brain functional plasticity associated with the emergence of expertise in extreme language control. *Neuroimage*, **114**, 264–274. https://doi.org/10.1016/j.neuroimage.2015.03.072

Holle H and Gunter TC (2007) The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, **19**, 1175–1192. https://doi.org/10.1162/jocn.2007.19.7.1175

Hostetter AB (2011) When do gestures communicate? A meta-analysis. *Psychological Bulletin*, **137**, 297–315. https://doi.org/10.1037/a0022128

ISO. (2016a) Simultaneous interpreting – Mobile booths – Requirements (ISO Standard no. 4043:2016). Retrieved from https://www.iso.org/standard/67066.html

ISO. (2016b) Simultaneous interpreting – Permanent booths – Requirements (ISO Standard no. 2603:2016). Retrieved from https://www.iso.org/standard/67065.html

ISO. (2017) Simultaneous interpreting – Quality and transmission of sound and image input – Requirements (ISO Standard no. 20108:2017). Retrieved from https://www.iso.org/standard/67062.html

Jesse A, Vrignaud N, Cohen MM and Massaro DW (2000) The processing of information from multiple sources in simultaneous interpreting. *Interpreting*, **5**, 95–115. https://doi.org/10.1075/intp.5.2.04jes

Kelly SD (2017) Exploring the boundaries of gesture-speech integration during language comprehension. In Church RB, Alibali MW and Kelly SD (eds), *Why gesture?: How the hands function in speaking, thinking and communicating*. Amsterdam, Netherlands: John Benjamins Publishing Company, pp. 243–265. https://doi.org/10.1075/gs.7.12kel

Kelly SD, Creigh P and Bartolotti J (2010) Integrating speech and iconic gestures in a Stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience*, **22**, 683–694. https://doi.org/10.1162/jocn.2009.21254

Kelly SD, Healey M, Özyürek A and Holler J (2015) The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*, **22**, 517–523. https://doi.org/10.3758/s13423-014-0681-7

Kelly SD, Özyürek A and Maris E (2009) Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, **21**, 260–267. https://doi.org/10.1177/0956797609357327

Kendon A (1995) Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, **23**, 247–279. https://doi.org/https://doi.org/10.1016/0378-2166(94)00037-F

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

Langton SRH and Bruce V (2000) You must see the point: Automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 747–757. https://doi.org/10.1037/0096-1523.26.2.747

Marian V, Blumenfeld HK and Kaushanskaya M (2007) The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, **50**, 940–967. https://doi.org/doi:10.1044/1092-4388(2007/067)

Mayberry RI and Jaques J (2000) Gesture production during stuttered speech: Insights into the nature of gesture-speech integration. *Language and gesture*, 199–214.

Mayberry RI and Nicoladis E (2000) Gesture reflects language development: Evidence from bilingual children. *Current Directions in Psychological Science*, **9**, 192–196. https://doi.org/10.1111/1467-8721.00092

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

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

McNeill D (2005) *Gesture and thought*. Chicago, IL, US: University of Chicago Press. https://doi.org/10.7208/chicago/9780226514642.001.0001

Moser-Mercer B (2000) Simultaneous interpreting: Cognitive potential and limitations. *Interpreting*, **5**, 83–94. https://doi.org/https://doi.org/10.1075/intp.5.2.03mos

Rayner K (1998) Eye movements in reading and information processing: 29 years of research. *Psychological Bulletin* **124**, 372–422.

R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Rennert S (2008) Visual input in simultaneous interpreting. *Meta*, **53**, 204–217. https://doi.org/https://doi.org/10.7202/017983ar

Rose ML (2006) The utility of arm and hand gestures in the treatment of aphasia. *Advances in Speech Language Pathology*, **8**, 92–109. https://doi.org/10.1080/14417040600657948

Seeber KG (2011) *Multimodal input in simultaneous interpreting. An eye-tracking experiment*. Paper presented at the 1st International Conference TRANSLATA, Translation & Interpreting Research: Yesterday - Today - Tomorrow, Innsbruck.

Seeber KG (2015) Cognitive load. In Pöchhacker F, Grbic N, Mead P and Setton R (eds), *Routledge encyclopedia of interpreting studies*. Oxford and New York: Routledge, pp. 60–61.

Seeber KG (2017) Multimodal processing in simultaneous interpreting. In Schwieter JW and Ferreira A (eds), *The handbook of translation and cognition*. New Jersey: Wiley Blackwell, pp. 461–475. https://doi.org/https://doi.org/10.1002/9781119241485.ch25

Snodgrass JG and Vanderwart M (1980) A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 174–215. https://doi.org/10.1037/0278-7393.6.2.174

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

Streeck J (2008) Gesture in political communication: A case study of the Democratic presidential candidates during the 2004 primary campaign. *Research on Language and Social Interaction*, **41**, 154–186. https://doi.org/10.1080/08351810802028662

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press

Strobach T, Becker M, Schubert T and Kühn S (2015) Better dual-task processing in simultaneous interpreters. *Frontiers in psychology*, **6**, 1590–1590. https://doi.org/10.3389/fpsyg.2015.01590

Sueyoshi A and Hardison DM (2005) The role of gestures and facial cues in second language listening comprehension. *Language Learning*, **55**, 661–699. https://doi.org/https://doi.org/10.1111/j.0023-8333.2005.00320.x

Szekely A, Jacobsen T, D'Amico S, Devescovi A, Andonova E, Herron D, Lu CC, Pechmann T, Pléh C, Wicha N, Federmeier K, Gerdjikova I, Gutierrez G, Hung D, Hsu J, Iyer G, Kohnert K, Mehotcheva T, Orozco-Figueroa A, Tzeng A, Tzeng O, Arévalo A, Vargha A, Butler AC, Buffington R and Bates E (2004) A new on-line resource for psycholinguistic studies. *Journal of memory and language*, **51**, 247–250. https://doi.org/10.1016/j.jml.2004.03.002

Tommola J and Lindholm J (1995) Experimental research on interpreting: Which dependent variable? In Tommola J (ed), *Topics in interpreting research*. Turku: Centre for Translation and Interpreting. University of Turku, pp. 121–133.

Wu YC and Coulson S (2007) Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*, **14**, 57–63.

Yudes C, Macizo P and Bajo T (2011) The influence of expertise in simultaneous interpreting on non-verbal executive processes. *Frontiers in psychology*, **2**, 309–309. https://doi.org/10.3389/fpsyg.2011.00309

https://doi.org/10.1017/S136672892200058X Published online by Cambridge University Press