Parallèles 37(1) April 2025

Modeling gestural alignment in spoken simultaneous interpreting: The role of gesture types

Inés Olza

Institute for Culture and Society. University of Navarra

Abstract

This article explores gestural alignment in spoken simultaneous interpreting, analyzing whether and how the interpreters under scrutiny align with the gestural behavior of a visible speaker-source, and which gesture types by the speaker-source more often prompt a gesturally aligned response by the interpreters. The paper offers a mixed-methods analysis of a set of multimodal data collected under (quasi-)experimental conditions in a real court interpreting setting during spoken training exercises performed by two novice interpreters. This study relies on the findings of a previous exploratory approach to the same dataset (Olza, 2024), where different degrees of gestural alignment were found and defined. In this study, the variable *gesture type* is used to systematically examine a new sub-sample of the same data and to compare the performance of the two novice interpreters. Results show that iconic gestures elicit higher degrees of alignment by both interpreters. The findings are not conclusive, though, when relating the (non-)representational nature of gestures by the speaker-source, nor their (non-) semantic value, to the degree of replication of such gestures by the two interpreters. Future research will rely on broader datasets obtained from more experienced interpreters engaged in tasks that more accurately reflect their actual practice.

Keywords

Gesture, alignment, spoken simultaneous interpreting, multimodal data, gesture types

 DOI:10.17462/para.2025.01.05

1. Introduction

1.1. Gesture in spoken simultaneous interpreting: previous studies

This paper aims to contribute to the growing body of research on gesture in spoken and signed-to-spoken simultaneous and consecutive interpreting, where a 'multimodal turn' in training, practice and analytical approaches is proposed, thanks mainly to the pervasive presence and use of video in professional and academic settings (Salaets & Brône, 2020). Indeed, the possibility, for the interpreters, of fully accessing and watching the speaker-source's performance (including non-verbal behavior) and, for the researchers, of recording and scrutinizing both the speaker-source's and the interpreter's multimodal behavior, makes it possible to conduct in-depth systematic analyses of gesture in interpreting tasks (Chwalczuk, 2021; Stachowiak-Szymczak, 2019; Zagar Galvão, 2015), and to relate the gestural behavior of the interpreters to that of the speaker-source (Olza, 2024; Zagar Galvão, 2013).

Specifically, empirical research on gesture within the subfield of spoken simultaneous interpreting[1] has focused thus far on three complementary strands.

(a) Studies aiming to define the presence and role of gesture within the overall performance of professional interpreters and/or trainees, mainly to determine its role in processing the cognitive load demanded by interpreting tasks in experimental (Stachowiak-Szymczak, 2019, chapters 5 and 6) and naturalistic settings (Chwalczuk, 2021; Fernández Santana & Martín de León, 2022; Iriskhanova et al., 2023; Martín de León & Fernández Santana, 2021; Zagar Galvão, 2015, 2020), with elicited or real interpreting tasks.

(b) Research integrating the advances in multimodal (interaction) studies into the analysis of the interpreters' performance, including fine-grained analyses of the deployment of certain gesture types in their behavior (e.g. gaze and beat gestures, in Stachowiak-Szymczak, 2019; iconic gestures, in Fernández Santana & Martín de León, 2022; metaphoric gestures, in Leonteva et al., 2023); the role of gesture in managing interpreting disfluencies (Cienki, 2024); or the complex interactional dynamics shaping gesture and language in interpreting and interpreter-mediated contexts (Krystallidou, 2020), among other perspectives.

(c) Incipient research on the degree of gestural convergence between the speaker-source and the interpreter, with several qualitative and mixed-methods analyses providing preliminary empirical evidence of how interpreters often reproduce in their own discourse the gestures they observe in the speaker-source (Chwalczuk, 2021; Janzen et al., this special issue; Leonteva et al., 2023; Zagar Galvão, 2013;). Within this strand of research, especially iconic and metaphoric gestures were analyzed across the speaker-source and the interpreters' performance (Chwalczuk, 2021, section 4.1.1; Leonteva et al., 2023).

Within this background, our study aims to add to the understanding of the cognitive mechanisms that regulate spoken simultaneous interpreting, with a focus on how the gestural convergence exhibited by the interpreters with regard to the speaker-source may be described and discussed from the tenets of both alignment theories (section 1.2) and gesture studies (namely, those leading to the definition and characterization of gesture types; section 2.3.1).

[1] A recent overview of the research on gesture in other interpreting modalities (signed-to-spoken interpreting) and types (consecutive and distance interpreting), as well as in spoken simultaneous interpreting, was offered in the panel 'Gesture in spoken and signed-to-spoken language interpreting', convened by Sílvia Gabarró-López and Alan Cienki at the 18th International Pragmatics Conference (2023 IPrA Conference, Université libre de Bruxelles, July 2023), where around 20 scholars in pragmatics, cognitive and applied linguistics discussed the latest advancements in gesture analysis in interpreting. A first approach to the dataset of this study was presented at this panel. I would like to thank the convenors and participants for their insightful comments and suggestions.

1.2. Cognitive and gestural alignment: previous approaches and key definitions

In this paper, gesture is viewed as both a reflection and a shaper of human thought, serving as a material anchor to explore the embodied mental mechanisms that underlie language use (Cienki, 2022; McNeill, 1992, 2005). One of these large-scale cognitive operations that unfold across diverse kinds of human behavior, including language, is alignment. From an external perspective that observes a given subject's behavior, alignment encompasses the dynamics of convergence and divergence of his/her actions relative to others. These behavioral 'movements' materialize in changes and adaptations (accommodation) of his/her communicative behavior at different levels (verbal, paraverbal, non-verbal) (Giles & Ogay, 2007, p. 295). In other words, while engaging in linguistic interaction, speakers monitor their own behavior and that of their interlocutors, and consequently —even 'strategically'—approach (align with) or distance (misalign) their behavior relative to that of others. This adaptation may occur at a local level, through alignment in specific linguistic and gestural choices, or unfold in a sustained and progressive manner throughout an entire conversational exchange (Fusaroli & Tylén, 2016).

Such a definition of alignment was first proposed within the framework of Communication Accommodation Theory (CAT), rooted in social psychology and sociolinguistics (Giles et al., 1991; Giles & Ogay, 2007). It was later revisited and expanded by cognitive and behavioral studies, which have shown that (mis)alignment regulates not only face-to-face interaction and communication but also all kinds of human behavior involving cooperation between individuals. This includes joint actions ranging from physical manipulation of objects (e.g., cooking together) to symbolic tasks (e.g., playing together). Within this cognitive and behavioral framework, alignment has been defined from two complementary perspectives.

First, it has been described and explained as a material manifestation of the wider *priming* principle that regulates human interaction, taken as "an automatic, bidirectional process operating in parallel on several different levels of representation" (Healey, 2004, p. 201), through which the interacting individuals —the interlocutors, in the case of communication— couple their respective situational models, that is, their mental representations of the situation and/or issues under discussion (Pickering & Garrod, 2004, sections 2.1-2.3). As a result, interlocutors not only cooperate during interaction but also align through a form of 'mimetic' behavior, where they converge by 'imitating' each other's actions. Going beyond the logics of stimuli-response underlying the described priming views (Doyle & Franck, 2016; Krauss & Pardo, 2004), the second big approach to alignment proposes to analyze it under the scope of *grounding* and interpersonal synergy (Fusaroli & Tylén, 2016), as a form of synchronized activity which is negotiated in a relational way, with wider room for the joint attention and cooperative action (Eilan et al., 2005; Goodwin, 2018) that characterize any form of human communicative exchanges. Here, alignment strongly relies on the common goals and common ground, and the communicative dynamics established between the interlocutors in concrete, genre-based, and situated interaction, in a similar way to how conversational analytic approaches describe them (Riordan et al., 2014; Stivers, 2008; Stivers et al., 2011). All in all, both approaches stress that the participants engaged in communicative interactions tend to coordinate and converge, i.e., align in their behavior, exhibiting various degrees of mutual 'mimesis' at all linguistic levels (phonetic, lexical, syntactic, semantic), including the gestural one, which has remained largely unaddressed in alignment and accommodation studies until recently (Bergmann & Kopp, 2012; Kimbara, 2006; Kopp & Bergmann, 2013; Rasenberg et al., 2020), most of these recent studies focusing on data elicited and collected in laboratory settings. While contributing to bridge the gap in the study of gestural alignment in interaction, this article aims to address it through the analysis of ecologically valid data of live exercises by novice interpreters conducted in a real, naturalistic setting.

Furthermore, this study relies on the existing body of research on the 'interactive' nature of simultaneous interpreting, that is, the special cognitive and behavioral relationship between the speaker-source, the interpreter, and the recipient of the interpreter's performance, which is claimed to affect the gestural behavior of the interpreter (Chwalczuk, 2021, section 4.1.1; Janzen et al., this special issue; Leonteva et al., 2023). In line with these studies, we assume that interpreters align not only with regard to the speaker-source but also towards the recipients of their performance. In some cases, this could explain why they do not fully align with the speaker-source's gestural behavior, as other types of gestures might be better understood by their audience, as shown, for instance, by Janzen et al. (this special issue). Also, another obvious fact should be noted: even if interpreters seek to 'maximally align' with the speaker-source and his/her behavior and frame of understanding, they do *not* actually *interact* with him/her, at least in the sense that prevails in accommodation and alignment studies, where 'regular' communicative exchanges, that is, those with a dynamic exchange of speaker-listener roles between the interlocutors are examined. This would also explain why interpreters often reproduce 'self-adapted' versions of the gestures carried out by the speaker-source; for instance, simplified gestures that match better with time and cognitive constraints (Leonteva et al., 2023), or gestures that blend their own perspective with that of the speaker-source (Janzen et al., this special issue). Accordingly, although this study mainly focuses on the role of gesture types in modeling gestural alignment in simultaneous interpreting, these distinctive features of the interpreting tasks will also be taken into account in the discussion of results (see section 4).

2. Study design

This study relies on the findings of a previous exploratory approach (Olza, 2024) to the same dataset that is examined here (see section 2.1). In this first study, we conducted a mixed-methods analysis of spoken interpreting data audiovisually recorded in a natural professional setting (courtroom). The study quantified in a basic descriptive way, qualified and compared the degree of gestural alignment towards the same speaker-source exhibited by two distinct novice interpreters, who were recorded while working at the same time in the mentioned setting. The results of this previous research included a taxonomy of the different degrees of gestural alignment found in the data (see also section 2.3), with a good number of instances where the observed interpreters actually mimicked the speaker-source's gestures in type, form and function. Finally, the data analyzed in Olza (2024) were also categorized in an exploratory manner according to several basic gesture types (iconic and metaphoric gestures; discourse-structuring gestures; gestures for modality and stance), which allowed to formulate hypotheses on the higher or lower tendency of certain gesture types to be replicated by the interpreters. In the present paper, these hypotheses on the influence of gesture types on gestural alignment are retaken, expanded and tested in a more granular and systematic way. In sections 2.1 and 2.2, the main features of the study design are presented against the backdrop of the analysis conducted in Olza (2024), so as to explain how the present study advances the understanding of the gesture types that more often prompt an aligned response by the simultaneous interpreters in our data.

2.1. Data

The multimodal data examined here and in Olza (2024) were obtained at real training sessions for novice legal interpreters organized by the interpretation directorate of an official

international court[2]. The interpreters were postgraduate fellows immersed in a specialized training program aimed at integrating them into legal interpreting booths in international institutions. A complete approx. 30 min training session was recorded. This consisted of a live interpreting exercise carried out in a real medium-sized courtroom, where the speaker-source (male) sat at the main orator's position (central front) and the trainees (four subjects) occupied separate booths in both sides of the room. In addition to the trainers (experienced interpreters) of the four novice interpreters, who were sitting next to them in their respective booths, there was no external audience in the room. Due to equipment limitations, only two trainees were recorded.

It should be noted that the speaker-source delivered a speech in Spanish on non-legal issues related to the history of technology. In fact, although they were held by and for court interpreters, the main speech in the training sessions at this particular institution did not necessarily deal with legal issues —it could describe or explain any kind of issue, as happens in our data, where the speaker-source exposed the history of the Thermomix and the dishwasher. This type of non-specialized exercise was usually conducted in the first stages of the training program. In our data, two novice interpreters were recorded: Interpreter 1 (female) worked from Spanish into spoken English; and Interpreter 2 (female) worked from Spanish into spoken French. Visual access to the speaker was similar for both interpreters, as shown below in Figure 1. Three cameras recorded the training and were situated respectively at the right of the main speaker, and directed at both recorded interpreters. The cameras did not interfere with nor block the activity and visual access of the participants. In fact, it is important to note that, during the exercise, the interpreters directed their gaze towards the speaker-source most of the time. The only times they did not look at him were when they appeared to be writing down dates, numbers, and proper names on the papers in front of them. This direction of their gaze thus reinforces the hypothesis that their gestural behavior was aligned with that of the speaker, and was not a result of chance, for example.

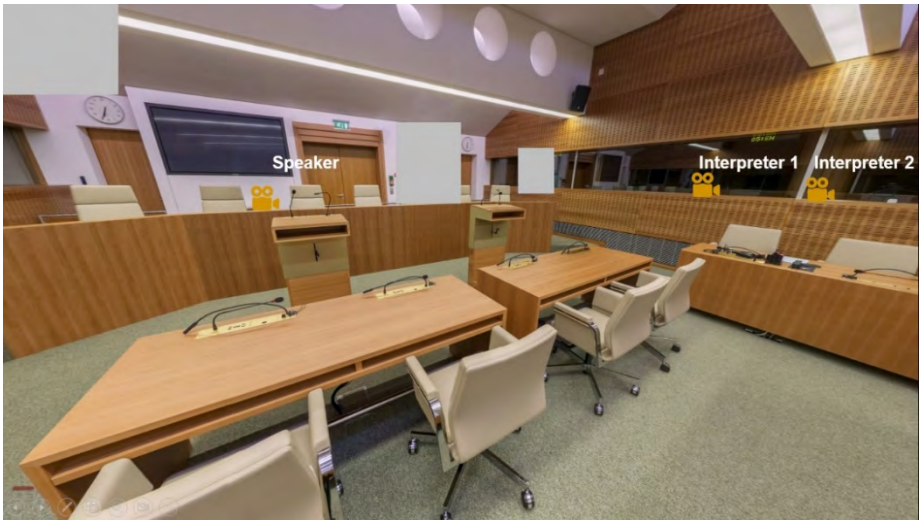


Figure 1. Recording setting (real courtroom): speaker-source, interpreters, and video-cameras

[2] The name and coordinates of the institution are not facilitated due to EU regulation on data protection. Before the sessions, the researcher in charge of the study presented it to the participants, who were able to ask any questions they had before signing the corresponding informed consent form. Previously, the study had received approval from the Research Ethics Committee of the University of Navarra (approval certificate nr. 2017.021).

2.2. Research questions and hypotheses

The analysis was guided by the following research questions and hypotheses, which emerge from the state-of-art described in section 1 and seek to improve and expand the results obtained in our previous study (Olza, 2024).

- **Research question 1.** Do the different gesture types by the speaker-source prompt diverse degrees of alignment by the interpreters in our data?

  Hypothesis 1. *Gestures that do not relate to the speech content (beats, self-adaptors) prompt lower degrees of gestural alignment by the interpreters.*

  The hypothesis is supported by the view of interpreting tasks as discourse (speech) oriented activities, where a common ground of understanding is negotiated with the speaker-source and the recipients of such tasks. Therefore, it seems reasonable to think that gestures not relating to the speech content and structure, and more dependent on the individual style of the speaker-source, will be less often replicated by the interpreters.

  Hypothesis 2. *Within the realm of gestures relating to the representational (referential) or pragmatic (metadiscursive) meaning of speech, iconic gestures and discourse structuring gestures prompt higher degrees of gestural alignment.*

  This hypothesis relies on the results of recent studies that have preliminarily suggested that iconic gestures are very often mirrored by interpreters (Chwalczuk, 2021, section 4.1.1; Olza, 2024), followed by gestures with discourse structuring functions (deictic gestures pointing to discourse referents) (Olza, 2024). In the latter (Olza, 2024), the initial hypothesis was that only representational gestures (iconic and metaphoric) would prompt higher degrees of alignment by the interpreters, as they relate to the referential and conceptual content of discourse. However, in this previous study, discourse structuring gestures were more often replicated than metaphoric gestures. Therefore, relying on a different sub-sample within the same dataset, the present study aims to test and, possibly, replicate the results obtained in Olza (2024).

  Hypothesis 3. *Compared to other types of representational gestures, metaphoric gestures prompt lower degrees of gestural alignment.*

  The hypothesis accords with the unexpected results of our previous study (Olza, 2024), which may replicate here, and those by Leonteva et al. (2023), who show that, due to the cognitive load and time pressure of the tasks, interpreters tend to lower the cognitive complexity (e.g. mental imagery) of their gestural behavior as a response to metaphoric gestures by the speaker-source. In other words, they tend to 'simplify' their gestural output compared to that of the speaker.

- **Research question 2.** Does gestural alignment rely on individuals? Or, on the contrary, does it work similarly in both interpreters who were observed?

  Hypothesis 4. *The degree of gestural alignment exhibited by both interpreters is different due to personal styles and/or differences in fluency and performance quality.*

  The hypothesis emerges from the results of analyzing a different data subsample in Olza (2024), where Interpreter 1 and Interpreter 2 showed different degrees of gestural alignment towards the same speaker-source. The present study seeks to test these findings in a different subsample of the same dataset.

In summary, the research questions and hypotheses presented here aim, on the one hand, to replicate the main results obtained in Olza (2024), particularly to confirm or refute the differences observed in the interpreters' performance, and to test once again the notable

frequency with which iconic gestures were replicated by the novice interpreters. On the other hand, the present study makes novel progress in two directions: in a better systematization of the types of gestures analyzed, with the introduction of gestures unrelated to speech content (beats, which are numerous in the data; and self-adaptors); and in the introduction of the variable 'gesture related/unrelated to speech content' in the study design and discussion of results.

2.3. Methods

The three video recordings (Speaker-source, Interpreter 1, Interpreter 2) were analyzed and tagged separately using the annotation tool ELAN-6.5[3]. The analysis was run by a single coder (the author of this paper) according to the following steps.

2.3.1. Analyzing the speaker-source's gestural behavior: Sample and gesture types

Four 1-minute excerpts of the speaker-source's behavior were analyzed and later on used as a baseline for the comparative analysis of the performance of Interpreters 1 and 2. The excerpts were chosen randomly using an open-source aleatory choice generator[4], resulting in minutes 15:00-16:00, 17:00-18:00, 19:00-20:00 and 21:00-22:00[5]. The selected excerpts were qualitatively analyzed in ELAN. First, the presence of any gesture relevant to the speech content was annotated. The gestures relating to the speech content were temporally delimited and annotated using the tags [gesture type], [body part(s) involved], and [speech sequence going along with gesture].

The coding of gesture types becomes especially relevant for this study, as it revises and expands the gesture types coded in Olza (2024), where a first approach to the influence of gesture types by the speaker-source on the gestural performance of interpreters was offered. In Olza (2024), only representational and pragmatic gestures were distinguished and coded, relying on the following well-established categories (McNeill, 1992; Kendon, 2004): *iconic* —gestures exhibiting a close formal relationship to what is semantically conveyed in speech; *metaphoric* —gestures depicting a figurative image of an abstract concept; *discourse and information structure* —gestures pointing to the discourse referents/topics and/or relating to discourse structuring information; *modality and stance* —gestures for intensification or mitigation of the expressed content; and *gestures for negation*. Beats and self-adaptors (Ekman & Friesen, 1972) were excluded in this exploratory approach, as they do not relate to what is signaled or represented by speech. In this first study (Olza, 2024, section 4), the novice interpreters were observed to align more often with iconic gestures and gestures related to modality and stance (mainly, gestures for intensification of semantic properties conveyed in the speech sequence), with a notable degree of alignment observed also for gestures having discourse structuring functions (e.g. signaling enumerations). Metaphoric gestures by the speaker-source were the type less often replicated by the interpreters.

The results obtained in Olza (2024) had, however, several limitations. First, the complete gestural behavior of the speaker-source was not analyzed, as only gestures related to the speech content were tackled. As pointed out before, beat gestures and adaptors were not coded. Second, the category of gestures for modality and stance turned out to be more problematic than expected, as it involved a higher level of interpretation compared to the other categories, with formally diverse gestures to which the function of expressing the speaker's attitude was

[3] https://archive.mpi.nl/tla/elan (accessed on 15 August 2023).

[4] https://randomchoicegenerator.com/ (accessed on 23 February 2024).

[5] The four excerpts randomly selected and analyzed in Olza (2024) were different (2:00-3:00; 10:00-11:00; 20:00-21:00; and 27:00-28:00).

attributed. For example, the sweeping away gesture (palm down hands moving away in front of the speaker, metaphorically clearing his/her personal space; Bressem & Müller, 2014) was frequently interpreted as a modal gesture for intensification (meaning 'totally'), which resulted in extracting it from the category to which it originally belonged (metaphoric gesture).

To overcome these limitations, the present study analyzes *all* the gestures by the speaker in the selected excerpts, encompassing gestures related to the referential and pragmatic meaning of the speech component (Kendon, 2004, chapter 10), as well as those not relating to the speech content (beats and self-adaptors). Another advance with respect to Olza (2024) has to do with the final set of gesture types analyzed and annotated in the present study, whose definition relies on formal-functional criteria that seek to minimize interpretative biases. Thus, the behavior of the speaker-source was categorized according to the following gesture types (Ekman & Friesen, 1972; McNeill, 1992; Kendon, 2004).

Beats	Gestures that go along with the rhythmical pulsation of speech (i.e., prosody).
Deictic (discourse structure)	Gestures signaling or pointing to the discourse referents/topics.
Head shakes (negation)	Lateral head movements (prototypical gesture for negation).
Iconic	Gestures exhibiting a close formal relationship to what is semantically conveyed in speech.
Metaphoric	Gestures depicting a figurative image of an abstract concept.
Self-adaptors	Non-signaling gestures where one part of the body is applied to another part of the body, such as scratching one's head and face.

Table 1. Analysis of gestural behavior: gesture types

A basic descriptive quantification of the total number of gestures and gesture types that were identified within the speaker-source's excerpts is displayed below in Table 2. In total, 118 gestural units were identified and classified.

Speaker-source		
Gesture type	**Hits**	**Rate**
Deictic (discourse structure)	40	34%
Beat	24	20.3%
Iconic	24	20.3%
Metaphoric	24	20.3%
Adaptor	5	4.2%
Head shake (negation)	1	0.9%
Total	**118**	**100%**

Table 2. Speaker-source: total number of gestures and gesture types

2.3.2. Tracking the gestural response of the interpreters

In a second phase, we analyzed and annotated the interpreters' performance in the excerpts where they interpreted the speech uttered by the speaker-source in the minutes analyzed in the first phase (section 2.3.1). For Interpreter 1, minutes 19:09-20:09, 21:10-22:12, 23:10-24:10, and 25:10-26:10 were analyzed; for Interpreter 2, minutes 17:02-18:02, 19:04-20:05, 21:00-22:00, and 23:00-24:07 were examined. As mentioned above, the speaker-source's behavior was taken as a baseline to analyze the interpreters' performance. Therefore, the sequences

where the speaker-source gestured —sequence = verbal cue and relevant gesture going along with it— were contrasted with the corresponding interpretation by both interpreters. Their behavior in the corresponding sequences was qualitatively analyzed and annotated using the following tags.

Speech interpreted—same type of gesture	While interpreting the verbal-gestural sequence of the speaker-source, the interpreters perform the same kind of gesture as the speaker.
Speech interpreted—different type of gesture	While interpreting the verbal-gestural sequence of the speaker-source, the interpreters perform a gesture of a different type than that of the speaker-source.
Speech interpreted—no gesture	The interpreters translate the verbal sequence that is accompanied by a gesture in the speaker-source's performance, but they do not gesture themselves.
Speech not interpreted	The concrete speech sequence of the speaker-source is not interpreted by the professional, due to disfluencies or constraints in time and expertise.

Table 3. Analysis of interpreters' behavior: tags to define their degree of gestural alignment

These tags sought to track the overall degree of gestural alignment exhibited by both interpreters in response to the verbal and gestural cues observed in the speaker-source. Although it would indeed be relevant to incorporate them into future studies, further details such as the verbal behavior of the interpreters—the actual words used to interpret the verbal sequence under scope—or a thorough formal description of each gesture were not systematically coded, as the main aim of our analysis was to offer a comprehensive comparative approach of the degree of convergence in the gestural behavior of the speaker-source and the interpreters. All in all, as shown in Tables 2 (above, section 2.3.1) and 5 (below, section 3), a total of 270 gestural units were identified and classified in the study: 118 for the speaker-source, 72 for Interpreter 1, and 80 for Interpreter 2.

Returning to the coding methods, cases where the interpreter would perform a similar gesture as the speaker while not interpreting his discourse were not clearly found in our data. Instances where speech was not interpreted were due to disfluencies that did not allow the interpreter(s) to tackle the sequence at all. Consequently, they would simply skip to the next discourse chunk.

The tags included in Table 3 allowed to define a gradual typology of (non-)aligned behavior by the novice interpreters with respect to that of the speaker-source (see Figure 2). Examples of the different degrees of gestural alignment within the proposed continuum are included below (see Figures 3-5). In these examples, and in the remainder of the article, cases where the interpreters did not interpret the verbal sequence produced by the speaker will not be considered.

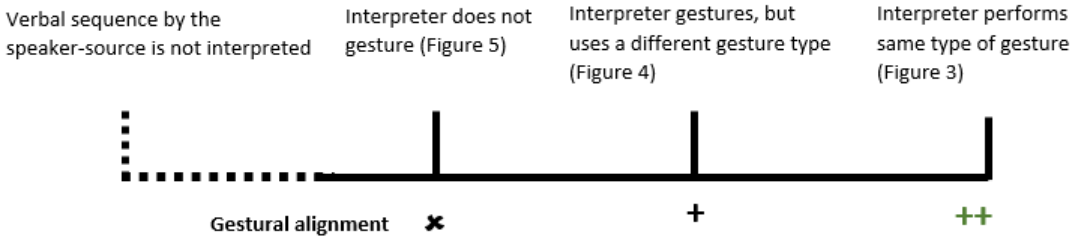


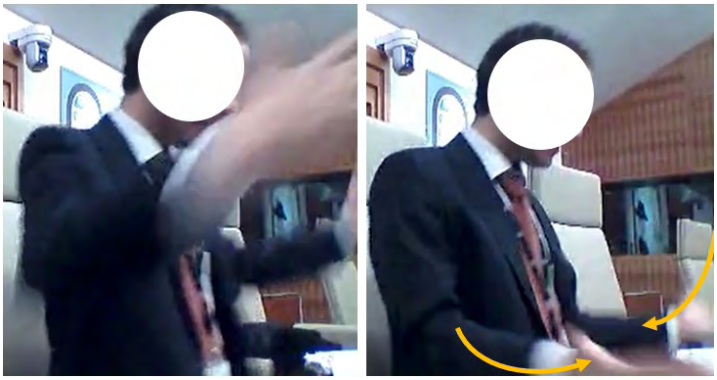
Figure 2. The 'gestural alignment continuum' (adapted from Olza, 2024)

Highest gestural alignment (++): Interpreter uses the same gesture type as the speaker-source

Speaker-source

Iconic gesture: depicting a round object with both hands.

(a) **(b)**



una caldera, es decir, un recipiente donde se calentaba agua
(a) **(b)**

a boiler, that is, a recipient where water has heated

Interpreter 2

Iconic gesture: depicting a round object with both hands.

(c) **(d)**

une chaudière à cuire
(c) **(d)**

a heater to cook

Figure 3. Interpreter 2 uses the same gesture type as the speaker-source

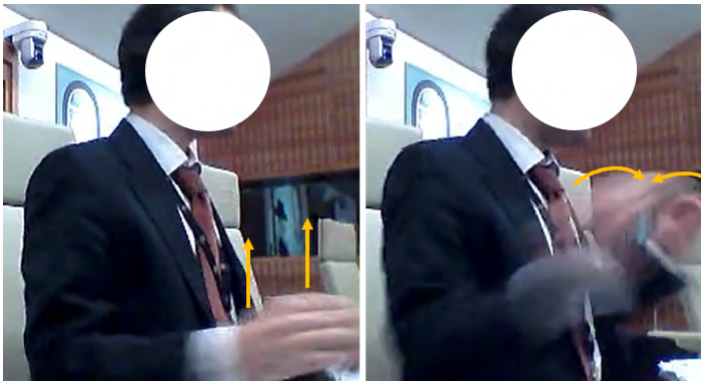
Incipient gestural alignment (+) — Interpreter gestures, but uses a gesture of a different type than that of the speaker

Speaker-source

Metaphoric gesture: first lifting, then moving both hands to the center, as if putting two things together.

The gesture is performed when mentioning the possibility of two persons taking care of the same activity.

(a) **(b)**



que uno baje la basura y otro se ocupe del lavaplatos
porque si los dos se ocupan del lavaplatos
(a) **(b)**
hay eh va a haber un problema

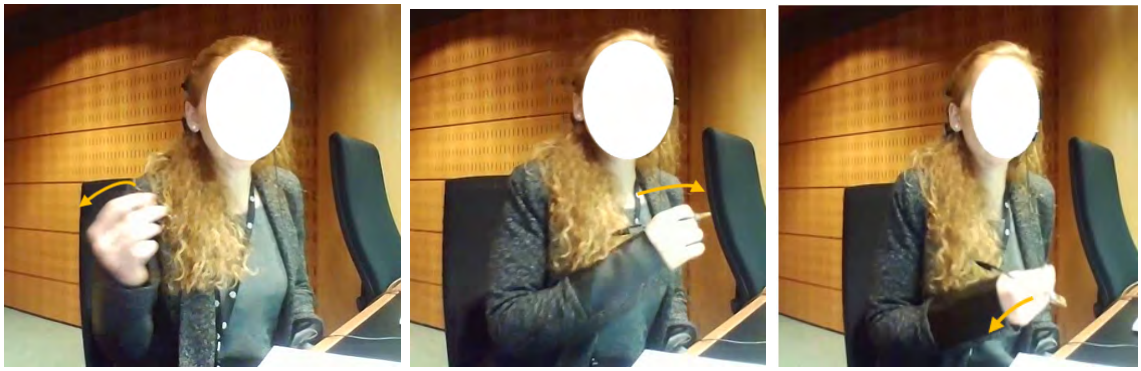
one should take out the trash, and the other should take care of the dishwasher, because if both take care of the dishwasher, there will be a problem

Interpreter 2

Deictic gesture (pointing forward to new referent, 'both') (caption e).

Previously, she used the same hand to point alternatively rightwards (caption c) and leftwards (d), when referring to the distribution of tasks between two subjects.

(c) **(d)** **(e)**



qu'un sorte les poubelles et que l'autre vide la vaisselle
(c) **(d)**
parce que si les deux s'occupent du lave vaisselle
(e)

one should take out the trash, and the other should take care of the dishwasher, because if both take care of the dishwasher

Figure 4. Interpreter 2 uses a gesture of a different type

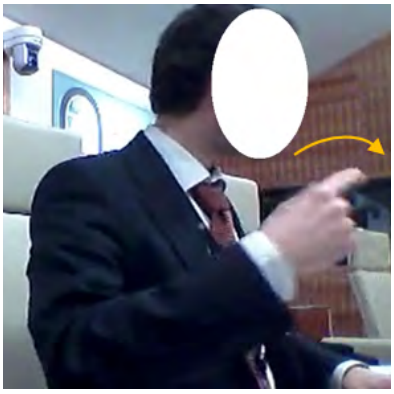
<table>
<tr><td colspan="2">No gestural alignment (×) — Speaker-source gestures but interpreter does not gesture</td></tr>
<tr><td>Speaker-source
Deictic gesture (pointing forward, new topic) with another person.
Previously, the same hand had pointed leftward, when saying treinta y seis años antes, ‘thirty-six years earlier’.</td><td>Interpreter 1
No gesture with another person.
Previously, her right hand points to her right when saying thirty-six years previous. After that, she holds the pen with both hands in a resting position and keeps them that way for the rest of the sequence.</td></tr>
<tr><td>(a)
  </td><td>(b)
  </td></tr>
<tr><td>treinta y seis años antes otra persona ya había patentado una máquina
(a)
thirty-six years earlier, another person had already patented a machine</td><td>thirty-six years previous to this another person had tried to patent a similar machine
(b)</td></tr>
</table>

Figure 5. Speaker-source gestures but Interpreter 1 does not gesture

3. Results

A descriptive quantification of the interpreters’ performance and a basic quantitative comparison of their behavior and that of the speaker-source are offered below in Tables 4 and 5.

	Interpreter 1		**Interpreter 2**	
	Hits	Rate	**Hits**	Rate
Speech interpreted—same type of gesture	50	42.4%	53	44.9%
Speech interpreted—different type of gesture	22	18.6%	27	22.9%
Speech interpreted—no gesture	28	23.7%	28	23.7%
Speech not interpreted	18	15.3%	10	8.5%
Speaker-source baseline →	**118**	**100%**	**118**	**100%**

Table 4. Overview of interpreters’ gestural performance

With regard to the interpreters’ overall performance (Table 4), two main tendencies should be noted. First, there are quite a few cases where the original speech of the speaker-source was not interpreted (15.3% of the cases for Interpreter 1; 8.5% for Interpreter 2). This is not surprising, as both are novice interpreters who sometimes experience disfluencies and miss or skip certain chunks of the speaker’s discourse. Interpreter 1 had more disfluencies or missed more speech sequences than Interpreter 2 (18 to 10), which might reveal an overall lower degree of interpreting competence.

At any rate, as second major tendency, there are no substantial differences in the degrees of gestural alignment exhibited by both interpreters in the sequences that were interpreted.

Furthermore, around 40% of the gestures by the speaker-source were mimicked by the interpreters through a gesture of the same kind.

	Baseline: gestures by speaker-source	Nr of hits where speech is interpreted along with *same type* of gesture				Nr of hits where speech is interpreted along with a gesture *of any type*			
		Interpreter 1		Interpreter 2		Interpreter 1		Interpreter 2	
Gesture type	**Hits**	**Hits**	**Rate***	**Hits**	**Rate***	**Hits**	**Rate***	**Hits**	**Rate***
Deictic (discourse structure)	40	13	32.5%	20	50%	20	50%	25	62.5%
Beat	24	10	41.6%	11	45.8%	15	62.5%	14	58.3%
Iconic	24	17	70.8%	14	58.3%	17	70.8%	22	91.6%
Metaphoric	24	9	37.5%	8	33.3%	16	66.6%	17	70.8%
Adaptor	5	0	0%	0	0%	3	60%	2	40%
Head shake (negation)	1	1	100%	0	0%	1	100%	0	0%
	118	**50**		**53**		**72**		**80**	

*Percentage of the speaker-source's gestures that prompt a gesture by the interpreter.

Table 5. Speech & gesture hits by the speaker-source that are interpreted along with a gesture

Table 5 reflects the gestural response of the interpreters according to gesture types by the speaker-source. As shown there, a total of 72 gestural responses by Interpreter 1 and 80 for Interpreter 2 were identified and classified. Looking at the hits by the speaker-source that receive a maximally aligned gestural response (gesture of the *same type*), iconic gestures clearly stand out as the type that more often elicit a mimicking response by the novice interpreters. It should also be noted that, although triggering a gestural response *of any type* by both interpreters in at least half of the cases (even in 60-70% in Interpreter 1), beats and metaphoric gestures get a maximally aligned response (*same gesture type*) in a smaller proportion, this reduction being much clearer for metaphoric gestures.

With regard to the gesture types that more often trigger *any kind* of gesture by the interpreters, iconic, metaphoric and beat gestures again stand out in frequency in the response by both interpreters. Moreover, iconic gestures trigger a very high gestural response by Interpreter 2 —in 91.6% of the cases, she gestures as a response to iconic gestures.

When comparing the performance of both novice interpreters, convergences and divergences arise at different levels, with a pattern that is not clearly identifiable. The greatest similarities are primarily observed in the frequency with which both interpreters respond to metaphoric and beat gestures, whether with a gesture of the same type or any other kind of gesture. The current sample size limits, though, the possibility of conducting statistical analyses of significant differences between the two interpreters. A substantial expansion of the analyzed data will allow for such a study in the future.

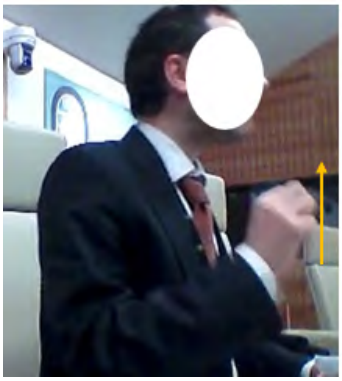
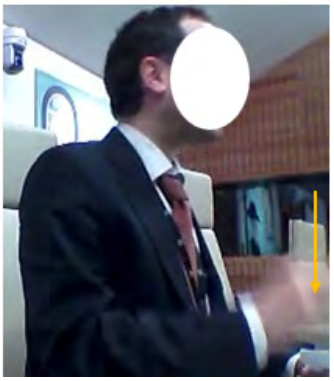
4. Discussion and conclusions

The results in section 3 allow to assess the research questions and hypotheses that were formulated above in section 2.2.

> **Research question 1.** Do the different gesture types by the speaker-source prompt diverse degrees of alignment by the interpreters in our data?
>
> > ***Hypothesis 1.*** *Gestures that do not relate to the speech content (beats, self-adaptors) prompt lower degrees of gestural alignment by the interpreters.*

The first hypothesis is not confirmed in our study, as it has been shown that sequences with beats are interpreted with gestures of any type in a good number of instances (62.5%, in Interpreter 1; 58.3%, in Interpreter 2), with still a notable proportion of cases where they are replicated by beat gestures (41.6%, in Interpreter 1; 45.8%, in Interpreter 2). Self-adaptors rarely appear in the sample analyzed in this paper and, consequently, their relationship to gestural alignment cannot be properly assessed.

However, the results obtained for beat gestures suggest that, in simultaneous interpreting, gestural alignment may not necessarily be driven by the distinction between semantic and non-semantic gestures (Ekman & Friesen, 1972; Kendon, 2004), that is, between gestures related or not related to the speech content with what is actually conveyed by words. In this vein, it could be claimed that simultaneous interpreting is of course guided by the semantic common ground that interpreters 'negotiate' with the speaker-source and the audience (*grounding views of alignment*), but *also* by other features of the linguistic behavior of the speaker-source, such as speech rhythm, prosody, and the beat movements that go along with them. Such a claim might —at least partly— support the *priming approaches to alignment* in a complementary and more comprehensive understanding of the coupling processes that regulate simultaneous interpreting. An example of an especially prominent alignment across all these aspects (rhythm, prosody, gesture) is provided in Figure 6. In this case, the speaker-source performs three beat gestures with his right hand when citing the title of a magazine section ('How to save your relationship'). These beats serve the function of parsing and stressing a segment of reproduced discourse (the section title). The title is cited verbatim, and so the hand also takes the form of a 'precision grip' gesture, as described by Kendon (2004, pp. 225-228). In Figure 6, the execution of the first of these beats by the speaker-source is visually depicted, with very broad and visible preparation and stroke phases. The interpreters' responses exhibit alignment on multiple levels: not only verbally, with a similar citation of the magazine section title, but also gesturally and in terms of rhythm and prosody, as the beat gestures they also perform with their right hands are synchronized with the same parts of the speech, emphasizing the quoted nature of the segment they accompany. Moreover, although both interpreters are holding a pen, the shape of their hands in some of their beats is compatible with a 'precision grip' gesture that serves the function of rhythmically parsing a segment of reproduced literal discourse.

Speaker-source	
en una sección de esta revista llamada **así salva usted** **beat 1 beat 2** su **relación** **beat 3** *in a section of this magazine called "how to save your relationship"*	**Beat 1 (preparation)**  **Beat 1 (stroke)** 

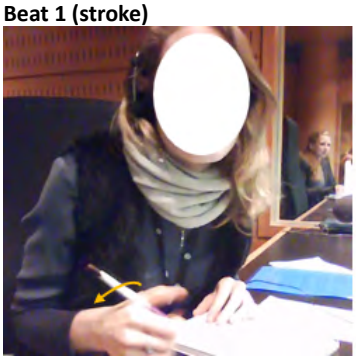
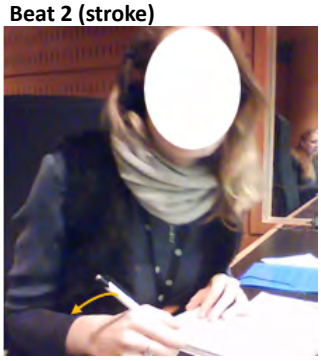
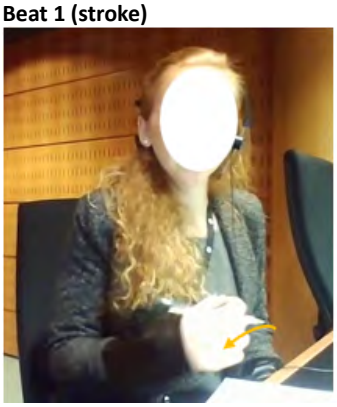
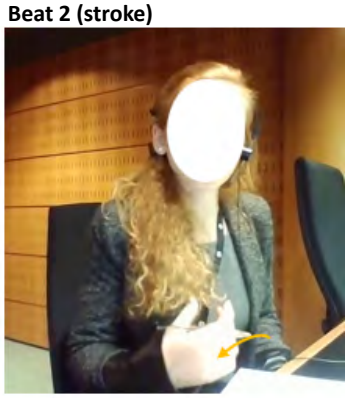
Interpreter 1	
in a section of the magazine how to **save** your re**lation**ship **beat 1** **beat 2**	**Beat 1 (stroke)**  **Beat 2 (stroke)** 
Interpreter 2	
dans la rubrique **comment sauver** sa **relation** **beat 1** **beat 2** **beat 3** *in the section "how to save your relationship"*	**Beat 1 (stroke)**  **Beat 2 (stroke)** 

Figure 6. Priming: alignment in gesture, rhythm and prosody

Hypothesis 2. *Within the realm of gestures relating to the representational (referential) or pragmatic (metadiscursive) meaning of speech, iconic gestures and discourse structuring gestures prompt higher degrees of gestural alignment.*

The hypothesis is confirmed only for iconic gestures, which clearly are the gesture type that is connected with a higher degree of gestural alignment on a more frequent basis and across both novice interpreters. This result confirms previous evidence in the same direction (Chwalczuk, 2021; Olza, 2024). As for deictic gestures with discourse structuring functions, results show a lower but notable triggering capacity for them, especially in Interpreter 2, who gestures in response to 62.5% of the cases, and replicates the same kind of gesture in 50% of the instances. Beat gestures seem to behave in a similar way to discourse structuring gestures, though. Therefore, our study is not conclusive on the operativity of the representational/ non-representational distinction (referential vs pragmatic gestures), nor the (non-)semantic one (beats and self-adaptors vs the rest of gesture types), to tackle gestural alignment in simultaneous interpreting.

Hypothesis 3. *Compared to other types of representational gestures, metaphoric gestures prompt lower degrees of gestural alignment.*

The hypothesis is not confirmed when the mere presence/absence of gesture by the interpreter is tracked, as metaphoric gestures follow iconic gestures in prompting a gestural response by the interpreters (66.6%, Interpreter 1; 70.8%, Interpreter 2). That being said, metaphoric gestures do exhibit more difficulties to elicit a maximally aligned response through another metaphoric gesture. The percentages reduce to 37.5% (Interpreter 1) and 33.3% (Interpreter 2) when looking at responses with the same type of gesture. As Leonteva et al. (2023, pp. 830-831)

claim, the production of metaphoric gestures involves the depiction and projection of actions in a physical domain (e.g., holding, molding, tracing, etc.) into an abstract domain, making it cognitively more demanding than performing non-metaphoric gestures, where only one representational domain is addressed. Regarding our results, and in line with these authors, it can be posited that, although metaphoric gestures may function as effective gesture primers, the demands of interpreting tasks make it difficult for interpreters to maintain the same level of metaphoricity in their gestures, leading them to use non-metaphoric gestures (e.g., iconic) in response to the speaker-source.

> **Research question 2.** Does gestural alignment rely on individuals? Or, by the contrary, does it work similarly in both interpreters who were observed?
>
> > ***Hypothesis 4.*** *The degree of gestural alignment exhibited by both interpreters is different due to personal styles and/or differences in fluency and performance quality.*

The results are inconclusive. On the one hand, differences for both interpreters were attested in fluency and competence (Table 4). Furthermore, the breakdown of their performance according to gesture types (Table 5) reveals some divergences in their gestural response the speaker source: in general terms, Interpreter 2 seems to respond more often to all kinds of gestures. In contrast, the two interpreters coincide in at least three main trends: they appear to be more sensitive to iconic gestures by the speaker-source; they also respond in notable ways to beats and metaphoric gestures; and they prefer other gesture types when aligning with metaphoric gestures. In our previous approach to another sample from the same dataset (Olza, 2024), clearer differences between the interpreters were observed. For instance, Interpreter 1 exhibited a much lower percentage of non-gestural hits –that is, of cases where the interpreted sequence was not accompanied by a gesture— compared to Interpreter 2 (7.3% for Interpreter 1; 25% for Interpreter 2). Although in the present study the performance of both interpreters was found to be more similar, a future analysis of the entire dataset will allow for proper statistical tests to better delineate the differences in their performance.

All in all, the most relevant findings of this study can be summarized in two directions. In the first place, the two interpreters under observation maximally aligned with the speaker-source at the gestural level, using the same type of gesture, in around 40% of the cases. Also, they gesturally responded to the speaker-source—irrespective gesture types—in more than 60% of the instances. To sum up: in our data, gestural alignment is more a norm than an exception. In the second place, iconic gestures were the gesture type that more often and better prompted gestural alignment by the interpreters. Beats and metaphoric gestures also elicited notable degrees of alignment.

A study like the one offered here shows that gestural alignment in simultaneous interpreting is still to be explored and understood in several uncharted territories. The results explained above nevertheless stress an uncontested claim in the field, which is that empirical evidence on interpreting tasks does not fit the ‘conduit model’, that is, it shows that interpreters do not merely transfer meanings from one language to another, mechanically decoding what the speaker says and then recoding it in exactly the same way in the target language, as described in the ‘conduit metaphor’ for language (Reddy, 1979, pp. 286-292), which was critically reviewed by Reddy himself in his seminal work (1979, pp. 297-310). Instead, their performance is better explained through a model that integrates the complex set of cognitive, linguistic, and behavioral conditions that influence the interpreters’ activity, which is more of a cooperative task than merely an ‘imitative’ one (Janzen et al., this special issue).

To further advance in the understanding of this complex set of factors and effects, a study like

this one will need to develop in several directions. For instance, the verbal response of the interpreters is still to be systematically studied, to analyze how the linguistic choices they make affect their own gestural behavior. In addition, a more thorough formal analysis of the gestures by both the speaker-source and the interpreters would allow to refine the conclusions offered here, as even the cases of what we have here considered as 'maximal gestural alignment' (same gesture type by the interpreters) exhibit interesting differences in the material articulation of the body movement, with different imagery and interpersonal features involved in them. Finally, other limitations of this study could be overcome with significantly broader data, as well as data even more closely aligned with the reality of professional interpreters. This would include gathering data from more experienced interpreters engaged in tasks that more accurately reflect their actual practice. As has been noted, the data for this study comes from training exercises with novice interpreters in a real courtroom setting, but in tasks different from strict legal interpretation. Therefore, it remains necessary to gather and analyze audiovisual data from experienced interpreters who either align or do not align gesturally with the speaker in real courtroom sessions.

In spite of its limitations, this study decidedly supports the call for a more multimodally oriented research on, and training of, simultaneous interpreters (Salaets & Brône, 2020). Videos and multimodal data are indeed the key to fully integrate the gestural dimension into the analysis of simultaneous interpreting. And this will, in turn, lead to a better awareness of the importance of multimodality in the interpreters' own professional performance.

5. Acknowledgements

This research was funded by the Spanish Ministry of Science and Innovation and FEDER/EU, grant numbers PGC2018-095703-B-I00 (MultiNeg) and PID2022-143052NB-I00 (MultiDeMe), and the Institute for Culture and Society, grant number ICS2020/09 (InMedio). Also, it was developed within the frame of the CoCoMInt Research Network of the Spanish Ministry of Science and Innovation (ref. RED2022-134123-T).

The author wants to thank the interpreters who generously agreed to be recorded for this study, and those who facilitated the fieldwork and recordings.

6. References

Bergmann, K. & Kopp, S. (2012). Gestural alignment in natural dialogue. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, *34*, 1326-1331. https://escholarship.org/uc/item/73z0q063 (accessed on 15 January 2024).

Bressem, J. & Müller, C. (2014). The family of away gestures: Negation, refusal, and negative assessment. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill & S. Tessendorf (Eds.), *Body—language—communication. An international handbook on multimodality in human interaction* (pp. 1592-1604). Mouton de Gruyter. https://doi.org/10.1515/9783110302028.1592

Chwalczuk, M. (2021). *La gestualité co-verbale en interprétation dans les services publics: Analyse contextualisée d'un corpus multimodal*. Doctoral dissertation, Université Paris Cité, France. https://theses.hal.science/tel-03509754/document.

Cienki, A. (2022). The study of gesture in cognitive linguistics: How it could inform and inspire other research in cognitive science. *WIREs Cognitive Science*, *13*(6), 1-17. https://doi.org/10.1002/wcs.1623

Cienki, A. (2025). Functions of gestures during disfluencies in simultaneous interpreting of speech, *Parallèles*, *37*(1), 29-46. https://doi.org/10.17462/PARA.2025.01.03

Doyle, G. & Franck, M. C. (2016). Investigating the sources of linguistic alignment in conversation. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 526-536). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-1050

Eilan, N., Hoerl, C., McCormack, T. & Roessler, J. (Eds.) (2005). *Joint attention: Communication and other minds*. Oxford University Press.

Ekman, P. & Friesen, W. V. (1972). Hand movements. *The Journal of Communication, 22*(4), 353-374. https://doi.org/10.1111/j.1460-2466.1972.tb00163.x

Fernández Santana, A. & Martín de León, C. (2022). Con la voz y las manos: Gestos icónicos en interpretación simultánea. *HERMĒNEUS,* 23, 225-271. https://doi.org/10.24197/her.23.2021.225-271

Fusaroli, R. & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science, 40*(1), 145-171. https://doi.org/10.1111/cogs.12251

Giles, H., Coupland, N. & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1-68). Cambridge University Press.

Giles, H. & Ogay, T. (2007). Communication Accommodation Theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293-310). Routledge.

Goodwin, C. (2018). *Co-operative action*. Cambridge University Press.

Healey, P. G. T. (2004). Dialogue in the degenerate case? *Behavioral and Brain Sciences, 27*(2), 201. https://doi.org/10.1017/S0140525X04330050.

Iriskhanova, O., Cienki, A. J., Tomskaya, M. V. & Nikolayeva, A. I. (2023). Silent, but salient: Gestures in simultaneous interpreting. *Research Result: Theoretical and Applied Linguistics, 9*(1), 99-114. https://doi.org/10.18413/2313-8912-2023-9-1-0-7

Janzen, T., Leeson, L., & Shaffer, B. Simultaneous interpreters' gestures as a window on conceptual alignment. *Parallèles*, *37*(1), 85-104. https://doi.org/10.17462/PARA.2025.01.06

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

Kimbara, I. (2006). On gestural mimicry. *Gesture, 6*(1), 36-61. https://doi.org/10.1075/gest.6.1.03kim

Kopp, S. & Bergmann, K. (2013). Automatic and strategic alignment of co-verbal gestures in dialogue. In I. Wachsmuth, J. de Ruiter, P. Jaecks & S. Kopp (Eds.), *Alignment in communication: Towards a new theory of communication* (pp. 87-108). John Benjamins.

Krauss, R. M. & Pardo, J. S. (2004). Is alignment always the result of automatic priming? *Behavioral and Brain Sciences*, *27*(2), 203-204. https://doi.org/10.1017/S0140525X0436005X

Krystallidou, D. (2020). Going video: Understanding interpreter-mediated clinical communication through the video lens. In H. Salaets & G. Brône (Eds.), *Linking up with video: Perspectives on interpreting practice and research* (pp. 81-202). John Benjamins.

Leonteva, A. V., Cienki, A. & Agafonova, O. V. (2023). Metaphoric gestures in simultaneous interpreting. *Russian Journal of Linguistics*, *27*(4), 820-842. https://doi.org/10.22363/2687-0088-36189

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

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

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

Olza, I. (2024). Gestural alignment in spoken simultaneous interpreting: A mixed-methods approach. *Languages, 9*(4), 151. https://doi.org/10.3390/languages9040151

Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169-190. https://doi.org/10.1017/S0140525X04000056

Rasenberg, M., Özyürek, A. & Dingemanse, M. (2020). Alignment in multimodal interaction: An integrative framework. *Cognitive Science*, *44*(11), e12911. https://doi.org/10.1111/cogs.12911

Reddy, M. J. (1979). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and thought* (pp. 284-324). Cambridge University Press.

Riordan, M. A., Kreuz, R. J. & Olney, A. M. (2014). Alignment is a function of conversational dynamics. *Journal of Language and Social Psychology*, *33*(5), 465-481. https://doi.org/10.1177/0261927X135123

Salaets, H. & Brône, G. (Eds.) (2020). *Linking up with video: Perspectives on interpreting practice and research*. John Benjamins.

Stachowiak-Szymczak, K. (2019). *Eye movements and gestures in simultaneous and consecutive interpreting*. Springer. https://doi.org/10.1007/978-3-030-19443-7

Stivers, T. (2008). Stance, alignment and affiliation during storytelling: When nodding is a token of preliminary affiliation. *Research on Language in Social Interaction*, *41*(1), 29-55. https://doi.org/10.1080/08351810701691123.

Stivers, T., Mondada, L. & Steensig, J. (2011). Knowledge, morality and affiliation in social interaction. In T. Stivers, L. Mondada & J. Steensing (Eds.), *The morality of knowledge in conversation* (pp. 3-24). Cambridge University Press.

Zagar Galvão, E. (2013). Hand gestures and speech production in the booth: Do simultaneous interpreters imitate the speaker? In C. Carapinha & I. Santos (Eds.), *Estudos de lingüística* (pp. 115-129). Coimbra University Press.

Zagar Galvão, E. (2015). *Gesture in simultaneous interpreting from English into European Portuguese: An exploratory study*. Unpublished Doctoral dissertation, University of Porto, Portugal.

Zagar Galvão, E. (2020). Gesture functions and gestural style in simultaneous interpreting. In H. Salaets & G. Brône (Eds.), *Linking Up with video: Perspectives on interpreting practice and research* (pp. 151-180). John Benjamins. https://doi.org/10.1075/btl.149.07gal

Inés Olza
University of Navarra
Institute for Culture and Society
Edificio Ismael Sánchez Bella
31009 Pamplona
Spain

iolzamor@unav.es

Biography: Inés Olza is Senior Researcher in Language and Cognition at the Institute for Culture and Society (ICS) of the University of Navarra, Spain. Within ICS, she is Principal Investigator of the Bonds, Creativity and Culture Research Group, and director of the Multimodal Pragmatics Lab (MuPra Lab). Her research focuses on figurative language, gesture, phraseology and interactional dynamics from the perspective of Pragmatics, Cognitive Linguistics and Multimodality.