ISSN 3041-217X (print)
ISSN 3041-2188 (online)
ALFRED NOBEL UNIVERSITY JOURNAL OF PHILOLOGY, 2026, 1(31)

UDC 81’25:81’33
DOI: https://doi.org/10.32342/anuJPh.2026.31.23

Multimodal Integration of Cognitive and Embodied Processes in Interpreting

Oleksandr Sodel
National University of Life and Environmental Sciences of Ukraine (Ukraine)

У статті розкривається мультимодальна інтеграція когнітивних і тілесних процесів при усному перекладі, як послідовному, так і синхронному, шляхом аналізу взаємодії вербальних, візуальних і кінетичних модальностей, зокрема того, як рухи очей, жести та зоровий контакт забезпечують конструювання смислу, регуляцію когнітивного навантаження й оптимізацію комунікативної ефективності у синхронному та послідовному перекладі. Проведено огляд теоретичних парадигм та емпіричних досліджень у сфері мультимодальності та суміжних комунікативних дисциплін, що заклало концептуальну основу розвідки. Здійснено системний аналіз продуктивності усних перекладачів із використанням інтегрованих методів відстеження погляду та визначення жестів, поєднуючи кількісні та якісні підходи для оцінки візуальної уваги та тілесної поведінки. Досліджено координацію візуальної уваги та моторних реакцій на різних фазах перекладацького процесу, що дозволило простежити механізми управління мультимодальними ресурсами в реальному часі. Виявлено закономірності синхронізації вербальних, візуальних і кінетичних каналів та оцінено їх когнітивно-прагматичні ефекти, включно з впливом на продуктивність, точність і комунікативну зрозумілість перекладу. На основі отриманих даних запропоновано інтегровану мультимодальну модель продуктивності усного перекладу, що поєднує когнітивну обробку інформації з тілесною взаємодією, забезпечуючи теоретичну базу та практичні рекомендації для підготовки фахівців із усного перекладу. У цьому дослідженні використовується змішаний експериментальний та спостережний *метод* із поєднанням кількісного аналізу даних зорового контакту та якісного аналізу жестів. Результати демонструють, що ефективна перекладацька діяльність базується на інтегрованій координації вербальних і невербальних каналів. Жести та погляд функціонують як зовнішні когнітивні інструменти, які сприяють організації інформаційних блоків, регуляції когнітивного навантаження та підвищенню продуктивності робочої пам’яті перекладача, забезпечуючи точність висловлювань в усному перекладі. Мультимодальна синхронізація виступає механізмом оптимізації поєднання когнітивних і тілесних ресурсів, акцентуючи роль мультимодальних стратегій у підвищенні професійної ефективності усних перекладачів.

Ключові слова: мультимодальність, ефективність усного перекладу, погляд, жести, зоровий контакт, когнітивне навантаження, тілесне пізнання, когнітивна обробка, мультимодальна модель процесу інтерпретації.

To cite this article: Sodel, O. (2026). Multimodal Integration of Cognitive and **Embodied** Processes in Interpreting. *Alfred Nobel University Journal of Philology*, 1 (31), 409-421, DOI: https://doi.org/10.32342/anuJPh.2026.31.23

In recent decades, interpreting has increasingly been viewed not merely as a linguistic activity but as a complex, multimodal form of human communication. While traditional studies in interpreting focused primarily on linguistic precision, cognitive effort, and memory use, modern approaches in applied linguistics and cognitive translation studies emphasize that interpreting naturally incorporates multiple modes of expression, such as speech, intonation, gaze, gestures, and body movements. These modes operate simultaneously to create and convey meaning between the interpreter, speaker, and audience.

Integrating multimodal analysis into interpreting research allows scholars to explore how visual and kinetic modalities contribute to the comprehension, processing, and transmission of messages, thereby offering a more holistic understanding of the interpreting process.

Maria Jesus Pinar highlights the integration of cognitive linguistics and multimodality, viewing meaning as co-constructed through multiple semiotic channels [Pinar, 2013]. She emphasizes how verbal, visual, and gestural cues interact with perception, memory, and attention, underlining the cognitive mechanisms that contribute to multimodal comprehension. Similarly, Anna Jelec explores multimodal patterns in cognition and communication, demonstrating that gestures, gaze, and posture systematically contribute to meaning-making [Jelec, 2020]. Her findings support Pinar's emphasis on cognitive foundations while providing empirical evidence for the synchronized use of modalities in natural communication.

Neil Cohn and Joost Schilperoord propose a cognitive framework in which multimodal language functions as an integrated system combining verbal and non-verbal modalities [Cohn, Schilperoord, 2024]. In contrast to Pinar, who emphasizes cognitive representations, they focus on the structural and systematic properties of multimodal communication, showing that gestures, gaze, and visual symbols act as linguistic elements, organizing meaning and facilitating both comprehension and production.

Similarly, Larysa Makaruk explores multimodal syntactic constructions in digital English communication [Makaruk, 2025]. Her research demonstrates that these multimodal constructions are systematically structured and can be classified by dominant semiotic components, revealing both the cognitive and pragmatic potential of integrating multiple modalities for effective message delivery. This work complements the cognitive and structural approaches of Pinar, Cohn and Schilperoord by emphasizing the interaction of verbal and non-verbal channels in practical communicative contexts, such as advertising and digital media.

Asli Özyürek explores multimodal language, from the perspective of interlinguistic interaction, showing that gestures and gaze are shaped by linguistic and cultural norms [Özyürek, 2021]. In contrast to Pinar and Cohn & Schilperoord, who focus on general cognitive mechanisms, Özyürek emphasizes differences across languages and interactional contexts, demonstrating that multimodal strategies reflect both universal cognitive processes and language-specific regularities – an insight especially relevant for studies in interpreting.

Kurt Feyaerts, Geert Brône, Bert Oben explore multimodality in interaction, emphasizing the social and pragmatic functions of gestures, gaze, and other bodily cues [Feyaerts, Brône, Oben, 2017]. Their approach complements cognitive aspects by illustrating how multimodal behavior is constructed in real time, supporting functions such as turn-taking, emphasis, and coordination between interlocutors. Unlike Pinar and Cohn, Feyaerts et al. offer a more dynamic and interaction-oriented view of multimodality.

These studies view multimodality as both cognitively and socially conditioned. Cognitive-linguistic approaches (Pinar, Cohn, Schilperoord, Jelec) emphasize mental representations and information-processing mechanisms, while interactionist approaches (Feyaerts et al., Özyürek) highlight real-time coordination and language- or culture-specific variability. In interpreting research, integrating these approaches allows for a comprehensive analysis of how interpreters use gaze, gestures, and speech to direct attention, support memory, and maintain interactional coherence across languages and contexts.

Research in interpreting often relies on Gile's Effort Model, which describes the management of cognitive load during listening, production, and memorization [Gile, 2021]. Developing this concept, Yana Boiko [Boiko, 2025] examines the manifestation of these cognitive efforts in business translation, highlighting how high terminology density, rapid speech rate, and cultural differences increase cognitive stress. Her research reveals adaptive techniques used by

interpreters, such as anticipation, segmentation, and reformulation, which facilitate information processing and maintain communicative precision in time-sensitive business negotiations.

Despite growing interest in multimodal communication, empirical research on interpreters' eye movements and gestures remains limited. While gesture studies have explored non-verbal behavior during public speaking and bilingual interactions, their application to professional interpreting remains underdeveloped. Similarly, although eye-tracking is widely used in reading and translation studies, it is rarely applied to the analysis of interpreting process. Therefore, systematic research is needed to explore how visual attention and gesture dynamics interact with cognitive processing during interpreting.

The *aim* of this study is to reveal the multimodal integration of cognitive and embodied processes in both simultaneous and consecutive interpreting, focusing on how mental operations involved in perception, comprehension, attention, working memory, and meaning reformulation are dynamically coordinated with bodily actions such as gaze behaviour, gestures, and posture during meaning construction, *cognitive load management, and interpreter-mediated communication.*

To achieve this aim, the following *tasks* have been formulated: 1) to analyze the interpreter's efficiency through the integration of eye-tracking and gesture analysis methods; 2) to examine how interpreters coordinate visual attention and bodily movements during different interpreting stages; 3) to identify patterns of multimodal synchronization and their cognitive-pragmatic implications for interpreting efficiency; 4) to propose a multimodal model of interpreting process that accounts for both cognitive processing and bodily interactional behaviour.

The study corpus comprised diverse materials intended to elicit naturalistic interpreting process while ensuring experimental control. It included conference speeches (5–7 minutes each) with formal and informational content, simulated professional dialogues reflecting business and academic contexts, and brief expository texts used for consecutive interpreting tasks.

This study uses *a mixed experimental and observational method*, integrating quantitative eye-tracking measures with qualitative gesture analysis to investigate interpreters' multimodal behavior. To compare multimodal strategies, two groups of 12 participants with varying levels of experience were recruited, including 6 experienced professional interpreters (group A) and 6 graduate student interpreters (group B). The groups were balanced to allow comparison of multimodal strategies across different levels of professional experience. All participants were fluent in both the source and target languages, ensuring that the observed differences reflected interpreting strategies rather than language proficiency.

The analysis involved systematic coding of multiple modalities using quantitative and qualitative approaches. Gaze data were analyzed for fixation duration, saccade direction, and shifts, providing a quantitative assessment of visual attention during the listening, reformulating, and presenting phases.

Gestures, classified as iconic, deictic, rhythmic, or metaphorical, were analyzed for frequency, function, and timing of speech to reveal how physical movements complemented verbal output. All data streams were time-aligned to allow cross-modal analyses, and multimodal correlations examined the relationships between gaze, gestures, and interpreting performance measures including accuracy, fluency, and cognitive load, offering a comprehensive view of how visual and kinetic modalities contribute to effective interpreting.

Multimodality implies the integration of several semiotic channels—verbal, visual, gestural, and prosodic—into the communication process. Modern linguistic and cognitive models emphasize that meaning is co-constructed through the interaction of these channels rather than solely through language. In interpreting, multimodal signals contribute to comprehension, attention management, and message delivery.

The multimodal approach allows researchers to study how interpreters coordinate different modalities to optimize interpreting process and manage cognitive load in real time. The interpreter acts as a mediator, dynamically coordinating verbal and non-verbal cues under time pressure. In addition to verbal information, interpreters use gestures, eye contact, and facial expressions as part of their communicative strategy to direct attention, ensure coherence of interaction, and enhance message clarity. The notion of the interpreter as a multimodal communicator aligns with theories of embodied cognition, which suggest that language processing and production are closely linked to sensory and kinetic processes. Thus, analyzing interpreters' multimodal behaviour provides

important insights into the cognitive mechanisms that facilitate message reformulation, emotional control, and audience engagement.

Contemporary research emphasizes the role of embodied cognition, suggesting that gestures and other bodily responses are not merely expressive but also actively support cognitive processes such as working memory, problem-solving, and attentional control. Thus, interpreters' multimodal behavior can be viewed as simultaneously communicative and cognitive, facilitating effective information processing in real time.

Embodied cognition theory posits that cognitive processes are grounded in sensorimotor experience rather than being purely abstract or symbolic [Macrine, Fugate, 2020; Gallagher, 2011]. Language comprehension, memory, and problem-solving depend on bodily states, gestures, and interactions with the environment. Shaun Gallagher emphasizes that cognition is inseparable from physical action and context, describing meaning-making as a bodily, interactive process.

In relation to interpreting, Jelena Milošević and Hanna Risku argue that interpreters act as bodily agents, whose gestures, posture, and gaze actively support cognitive functions such as attention allocation, memory retrieval, and real-time reformulation [Milošević, Risku, 2024]. Thus, the interpreter's body serves as a cognitive tool, enabling them to cope with the high cognitive loads inherent in both simultaneous and consecutive modes. Embodied cognition provides a theoretical basis for understanding multimodal behaviour in interpreting, where gestures, gaze, and prosodic cues integrate with verbal signal, enhancing meaning formation and listener comprehension. Experienced interpreters strategically coordinate these modalities to maintain fluency, ensure interactional coherence, and reduce cognitive strain [Milošević, Risku, 2024; Macrine, Fugate, 2020].

By incorporating embodied cognition, research in interpreting goes beyond purely linguistic or cognitive perspectives, integrating physical, perceptual, and social dimensions. The body mediates and facilitates cognitive processes, linking mental representations with external actions. For studies in interpreting, this emphasizes the importance of analyzing gestures, gaze, and other bodily cues as functional components of meaning-making and cognitive control.

Integrating embodied cognition with multimodal approaches enables the creation of a ***comprehensive model of interpreter performance***, demonstrating the interaction of verbal and non-verbal channels to ensure comprehension, manage cognitive load, and engage the audience. This theoretical perspective also supports the use of eye-tracking and gesture analysis in empirical studies, revealing the bodily strategies used by interpreters to optimize real-time information processing and enhance communicative effectiveness.

Gaze serves as a primary indicator of attention, focus, and cognitive processing, guiding comprehension and interaction management. Gestures, including iconic, deictic, metaphorical, and rhythmic gestures, facilitate verbal expression, memory retrieval, and emphasis, while prosody (intonation, stress, and rhythm) conveys discourse structure, emotional tone, and pragmatic intent. In interpreting, the coordinated integration of these modalities ensures accurate and coherent communication of meaning, especially when performing complex cognitive tasks in simultaneous interpreting.

Eye-tracking has become as a key method for studying attentional processes in interpreting. Ting Hu, Xinyu Wang and Haiming Xu review four decades of research, highlighting that eye-tracking provides insights into cognitive load, attentional allocation, and information processing strategies in both consecutive and simultaneous interpreting [Hu, Wang, Xu, 2022]. Jelena Vranjes and Geert Brône demonstrate its practical application in interpreter-mediated speech, showing how visual attention reflects strategies for comprehension, task management, and interaction monitoring [Vranjes, Brône, 2020]. Elisabet Tiselius and Kayle Sneed also show that gaze behavior in conversational interpreting reflects real-time bilingual information processing and coordination with interlocutors [Tiselius, Sneed, 2020].

Gestures play a functional role in supporting comprehension and communication. Melissa Singer, Joshua Radinsky, and Susan Goldman demonstrate that gestures facilitate meaning formation, memory retrieval, and verbal clarification [Singer, Radinsky, Goldman, 2008]. Bert Oben and Geert Brône reveal the close relationship between gaze and gesture, which enhances multimodal coherence [Oben, Brône, 2015]. Albert Giberga et al. emphasize that prosody and gestures jointly facilitate comprehension, particularly in pragmatic meaning-

making, demonstrating the synergistic action of multimodal cues in interpreting [Giberga et al., 2024].

Combining the results of eye-tracking and gesture studies reveals that visual and kinetic modalities are closely interconnected. Fixations often precede or coincide with gestures, reflecting anticipatory planning and cognitive unloading. Gaze patterns track attention to the speaker, visual aids, or notes, while gestures complement verbal information, maintaining coherence, pointing to referents, and highlighting key concepts.

Taken together, these studies highlight the importance of exploring multimodal strategies in interpreting. Eye-tracking and gesture analysis empirically confirm that interpreters coordinate visual attention and kinetic actions to manage cognitive load, enhance message clarity, and maintain interactional coherence. Understanding these dynamics is essential for both interpreter training and theoretical models of interpreting effectiveness, integrating cognitive, embodied, and interactive aspects.

Eye-tracking data analysis revealed distinct patterns of visual attention across the three stages of interpreting, highlighting both differences between and within the two groups of experienced professional interpreters (group A) and graduate student interpreters (group B) (Tab. 1).

Table 1

Comparison of Visual Attention Patterns Across Interpreting Stages

Interpreting Stage	Visual Attention Patterns	Group A	Group B	Cognitive/Functional Implications
Listening	Fixations on speaker's face and visual aids	More targeted fixations, moderate gaze shifts, some reliance on visual aids	Long fixations, frequent gaze shifts, reliance on visual cues	Group B exerts higher cognitive effort; Group A shows emerging efficiency
Reformulation	Fixations on critical information for retrieval	Intermediate fixation patterns, partially focused gaze	Diffuse, chaotic gaze patterns, high cognitive load	Group B struggles with attention allocation; Group A shows partial consolidation of visual strategies
Output	Fixations toward audience or notes	Inconsistent transitions, frequent returns to speaker or aids	More stable transitions than Group A, occasional backtracking	Group A shows ongoing effort and low automaticity; Group B demonstrates emerging automaticity and improved monitoring

While *listening stage*, the interpreters primarily focused their gaze on the speaker's face and available visual aids, reflecting attentional control and comprehension. Group B demonstrated longer fixations and frequent gaze shifts, indicating higher cognitive load and a greater reliance on visual cues. In contrast, Group A demonstrated more focused fixations with moderate gaze shifts, indicating increasing efficiency in extracting relevant information. These differences reflect variations in cognitive load management and selective attention, with Group B requiring additional effort to filter and prioritize incoming information.

During the *reformulation stage*, Group B exhibited diffuse and chaotic gaze patterns, indicating high cognitive load and less effective allocation of visual attention. Group A exhibited intermediate, partially focused fixations with occasional reliance on visual cues, indicating the development of multimodal strategies and more efficient cognitive control. These patterns suggest that experience facilitates more efficient information retrieval and organization, while less experienced interpreters rely heavily on visual control to compensate for less automated information processing.

During the *output stage*, fixations shifted toward the audience or the interpreter's notes, with the overall durations of fixations shortening compared to previous stages. Group A exhibited inconsistent transitions, frequently returning to the speaker or visual aids, reflecting ongoing cognitive effort and lower level of automaticity. Group B demonstrated more stable gaze transitions and smoother monitoring, though still not reaching the level of experienced professional interpreters. This stage reveals differences in anticipatory planning, multimodal integration, and

interactive monitoring, with Group B demonstrating emerging automaticity in the coordination of gaze with speech production.

To improve interpreting effectiveness, a number of targeted strategies can be used, each reinforced through training. **Gaze-focused training** aims to develop selective visual attention. For example, graduate student interpreters can watch short video clips of speakers and recognize key visual cues, such as gestures, slide highlights, or facial expressions, while ignoring irrelevant movements, thereby optimizing cognitive load management and prioritizing critical information.

Awareness and practice of **gestures** promote the purposeful use of iconic, deictic, and rhythmic gestures to improve memory and highlight essential points. Training can include interpreting short speeches, intentionally incorporating gestures that reflect the content, and then reviewing the recordings to assess the gestures' accuracy, rhythm, and coherence with speech.

Multitasking training incorporate listening, note-taking, and verbal inference to develop coordinated use of verbal and non-verbal channels. For instance, graduate student interpreters can interpret rapid speech by taking structured notes, gradually increasing the difficulty of tasks, or using controlled distractions to build cognitive resilience.

Feedback and reflective training utilize eye-tracking and video recordings, allowing graduate student interpreters to explore gaze patterns, fixation durations, and gesture synchronization. Comparison of the obtained data with expert models enables targeted adjustments to multimodal strategies and reinforces skills in effective attention management and gesture integration.

Finally, the **gradual increase in task complexity**—the gradually increase in the difficulty of interpreting training—promotes the development of automatic attentional control, cognitive management, and multimodal coordination. Tasks may range from short, clear speeches to dynamic, terminologically rich lectures and, ultimately, to interactive simulations with audience questions, developing the ability to effectively integrate verbal and non-verbal channels and developing adaptive, highly effective interpreters capable of handling real-time demands.

Gesture analysis revealed that both groups of experienced professional interpreters and graduate student interpreters used a range of gestures, each serving specific cognitive and communicative functions, with notable differences between Group A and Group B (Tab. 2).

Table 2

Comparison of Gesture Types and Functions

Gesture Type	Function	Group A	Group B	Interpretation
Iconic	Represent concrete concepts or actions	Used inconsistently, often to reinforce comprehension	Used strategically to emphasize key terms	Group B integrates gestures purposefully; Group A uses them compensatorily
Deictic	Direct attention or reference materials	Frequent; high reliance on visual cues	Less frequent; moderate use	Group A depends more on visual cues; Group B shows emerging independence
Metaphorical	Convey abstract ideas or conceptual relationships	Equally distributed	Equally distributed	Both groups use gestures for non-literal meaning similarly
Rhythmic	Mark speech rhythm, emphasize key points	Sporadically used	Consistently used; supports coherence and fluency	Group B shows more automatic integration; Group A uses gestures irregularly to support verbal output
Overall Function	Memory support, emphasis, communicative clarification	Less consistent; compensatory use	Greater consistency and integration with speech	Highlights strategic vs. compensatory multimodal behavior between groups

Iconic gestures, representing specific concepts or actions, were present in both groups; however, Group B used them more strategically to emphasize key terms, whereas Group A used them less consistently, often to visually reinforce understanding.

Deictic gestures, such as pointing or indicating objects, were more common in Group A, indicating a greater reliance on visual cues to direct attention and reference materials, while Group B used them more sparingly, reflecting emerging independence.

Metaphorical gestures, conveying abstract ideas or conceptual relationships, were equally distributed in both groups, suggesting that both groups relied on gestures to express non-literal meaning.

Rhythmic gestures, which establish a rhythm to speech and highlight key points, were used more frequently by Group B, promoting coherence and fluency, while Group A used them sporadically.

Overall, gestures in both groups of experiencd professional interpreters and graduate student interpreters facilitated memory retrieval, focused attention, and clarification of communicative points. Group B demonstrated greater consistency and automatic integration of gestures with speech, indicating more effective multimodal strategies and cognitive management. In contrast, Group A used gestures less regularly, often compensating for higher cognitive load and less automatic verbal processing. These differences are likely due to differences in experience, cognitive load management, and multimodal integration skills. Group B uses gestures strategically to enhance coherence and emphasize key points, while Group A uses gestures more selectively to facilitate comprehension and maintain clarity of interaction.

To improve the use of gestures in interpreting, several targeted strategies can be implemented through practical training. Graduate student interpreters can practice intentional, **iconic gestures** to enhance semantic clarity, for example, illustrating concepts like "growth" with upward hand movements or "separatiom" with separating gestures. Mirroring practice can help trainees align gestures with speech by observing their own gestures in real time and adjusting their rhythm and expressiveness.

Deictic gestures can be developed through combined gaze and pointing exercises. For instance, graduate student interpreters might interpret a short lecture by referring to specific objects, slides, or notes, intentionally directing their gaze and gestures to appropriate landmarks to enhance audience comprehension.

Metaphorical gestures can be integrated into consecutive interpreting practice to reinforce abstract or conceptual explanations. Training might involve interpreting passages with abstract ideas, such as "economic trends" or "organizational change," while practicing consistent and meaningful hand movements to represent the underlying concepts.

Rhythmic gestures can be synchronized with prosody through rhythm and tempo practice. Graduate student interpreters can read or interpret short texts matching hand movements to sentence stress, pauses, or accents. Video communication can help improve gesture synchronization and coordination with speech patterns.

Finally, a **progressive multimodal approach to training** combines gaze, gesture, and speech in increasingly complex simulated tasks. Training can begin with short, clear speeches, move to dynamic, information-rich lectures, and culminate in interactive situations with audience questions or unexpected interruptions. Incorporating reflection and feedback, such as viewing eye-tracking recordings and gesture videos, allows graduate student interpreters to consciously adjust their multimodal strategies. This step-by-step approach promotes the development of effective multimodal coordination, reduces cognitive load, and improves interactional coherence, preparing graduate student interpreters for working in high-pressure, real-time environments.

A multimodal analysis of experienced professional interpreters and graduate student interpreters revealed significant differences in the temporal coordination of gaze, gestures, and verbal information between members in Group A and B (Tab. 3).

Table 3

Comparative Analysis of Multimodal Coordination Patterns

Aspect	Group A	Group B
Gaze-Speech Coordination	Often delayed; fixations sometimes lag behind speech	Anticipates or complements speech segments; smoother transitions
Gesture Timing	Gestures frequently follow verbal output	Gestures frequently anticipate or align with speech
Prosody-Gesture Coordination	Less consistent coordination; rhythmic gestures not always timed with stress/ rhythm	Rhythmic gestures consistently coincide with prosodic emphasis, contributing to clarity
Integration of Multimodal Cues	Slower coordination between visual, kinetic, and verbal channels	Efficient integration, supporting cognitive load management
Overall Communication Effectiveness	Lower coherence; cognitive effort visible	Higher coherence; communication more fluid and structured

Group B demonstrated greater synchronization; gestures and gaze more often anticipated or complemented speech segments, indicating more efficient integration of multimodal cues and better management of cognitive load.

In contrast, Group A demonstrated less precise coordination: gestures often appeared after verbal responses, and gaze shifts were sometimes delayed, reflecting slower coordination between visual, kinetic, and verbal channels. Prosodic emphasis, including variations in intonation, stress, and rhythm, tended to more consistently match rhythmic gestures in Group B, contributing to the clarity and expressiveness of key points.

These results suggest that even among graduate student interpreters, greater experience facilitates more effective multimodal coordination, supporting both cognitive processing and coherent communication during interpreting.

Statistical analysis (Fig.1) revealed a significant relationship between gesture frequency and cognitive load measures, such as fixation duration and saccade count.

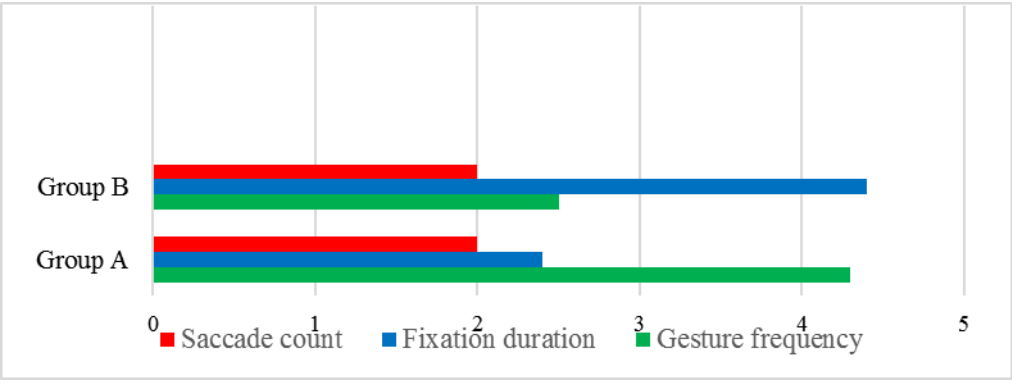

Fig. 1. *Statistical Analysis of Gesture and Cognitive Load Measures*

Group A demonstrated a more strategic use of gestures, coordinating gestures with verbal output to support memory retrieval and reduce cognitive effort, as reflected in shorter and more focused fixations.

In contrast, Group B demonstrated a higher but less systematic frequency of gestures, often using them inconsistently as a compensatory mechanism under higher cognitive load, accompanied by longer fixations and more diffuse gaze. These results suggest that experience influences how gestures are integrated into interpreting strategies to optimize cognitive processing and task performance.

The study's findings highlight the dynamic interplay between bodily actions and cognitive processes, including memory retrieval and attentional control. During interpreting, gestures often served as external memory aids; for example, subtle pointing movements or rhythmic hand movements helped interpreters structure information segments and recall terminological elements when paraphrasing. These gestures effectively reduced cognitive load, allowing working memory to focus on comprehension and retrieval rather than retention.

Gaze behavior also reflected complex attentional strategies. More experienced participants demonstrated proactive gaze patterns, anticipating important information, while less experienced—often exhibited reactive gaze, directing attention to notes or the speaker only after important material had been presented, sometimes resulting in delays in paraphrasing.

Furthermore, redirecting gaze toward the audience during the presentation indicated active monitoring of comprehension-related feedback. Brief eye contact allowed graduate student interpreters to maintain alignment and pragmatic coherence, particularly during information-dense segments.

Overall, the results confirm that bodily behavior not only expresses mental processes but also serves as **functionally integrated cognitive strategies**, enabling interpreters to regulate mental workload, maintain coherence, and improve overall interpreting efficiency and fluency.

Gestures and gaze play a central role in regulating the clarity, expressiveness, and engagement during the interpreting process. In particular, rhythmic gestures—small rhythmic hand movements synchronized with prosodic peaks—serve to *enhance intonation contours* and *highlight semantic foci.* For example, interpreters often used subtle downward or outward hand movements when rendering key terms or transitional phrases (e.g., *as a result, the main issue,* etc.), thereby drawing the listener's attention to important fragments of discourse.

Simultaneously, synchronized gaze shifts served as *visual markers of discourse segmentation*. When interpreters shifted their gaze toward the audience, emphasizing the main point, speech comprehension improved due to the alignment of verbal and visual cues. In contrast, brief gaze withdrawal (for example, looking down while formulating or recalling terms) corresponded to phases of internal processing, indicating the interpreter's cognitive load management.

Deictic gestures (such as pointing to slides, documents, or spatially imagined objects) and iconic gestures (illustrating size, direction, or movement) played an important role in *organizing discourse structure* and *denoting referents*. These multimodal cues not only helped the interpreter structure the flow of information but also helped the audience *follow complex content*, such as in technical or data-heavy presentations. Graduate student interpreters typically relied on spontaneous or compensatory gestures, often inconsistent with prosodic rhythm, which sometimes disrupted coherence of speech or diminished the communicative impact.

Gestures and gaze thus emerged as key multimodal resources, supporting both cognitive regulation (by externalizing thought processes) and interaction (by maintaining audience focus and interpretive coherence).

Preliminary results indicate clear differences in multimodal coordination during interpreting, depending on the mode.

In *simultaneous interpreting*, the combination of high cognitive load, time constraints, and parallel processing demands (listening, reformulating, and speaking in real time) necessitated stricter synchronization between modalities. Gestures in this mode were typically minimalistic and rhythmic—often limited to *short, rhythmic gestures* aligned with prosodic stress—so as not to disrupt auditory control or speech production. Gaze behavior was also limited: interpreters maintained *brief, strategic fixations* on the speaker or reference materials, while avoiding excessive audience engagement to maintain focus on the input.

In contrast, *consecutive interpreting* afforded graduate student interpreters greater temporal flexibility, allowing them to coordinate multimodal processes more purposefully. While taking notes and reformulating, graduate student interpreters demonstrated *greater gaze movements*, alternating between notes, the speaker, and the audience as well as a wider range of *iconic and deictic gestures*. These bodily resources facilitated discourse structuring, referent tracking, and listener engagement, effectively enhancing the clarity and expressiveness of speech after cognitive load was reduced.

Drawing on cognitive-pragmatic and embodied cognitive frameworks, this study proposes a **multimodal model of interpreting process** that conceptualizes interpreting as an integrated system of cognitive and bodily processes (Fig. 2).

In this model *gaze behavior* reflects attentional allocation and cognitive monitoring, demonstrating how interpreters distribute visual attention between the speaker, notes, and audience to regulate comprehension and convey information. *Gestures* act as external cognitive tools, functioning simultaneously as memory aids, communicative cues, and cognitive load regulators, thereby facilitating information retention and discourse organization. *Verbal output and prosody* represent the primary channels of meaning transmission, where rhythm, intonation, and stress interact with gestures and gaze to enhance clarity and coherence.

The model emphasizes the importance of temporal synchronization of these modalities—verbal, visual, and kinetic—and their dynamic, bidirectional contributions to both comprehension (processing the input text) and production (formulating the output text). It provides a conceptual framework for studying how experienced professional interpreters and graduate student interpreters coordinate multimodal resources to maintain accuracy, fluency, and pragmatic coherence throughout the interpreting process.

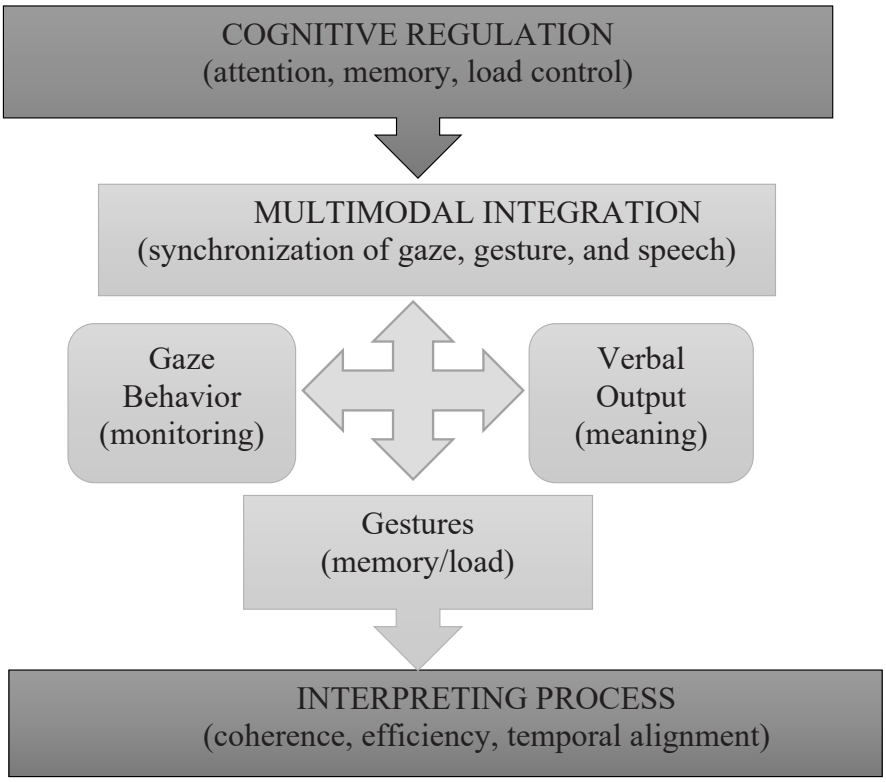

Fig. 2. *Multimodal Model of Interpreting Process*

Conclusions. This study demonstrated that interpreting is an inherently multimodal activity, in which gaze, gestures, and speech act as interconnected channels supporting comprehension, retention, and message transmission. The analysis revealed systematic patterns of visual attention, synchronized gestures, and precise timing of speech and emotional cues, which together facilitated cognitive load management and enhanced communicative clarity. Conversely, less coordinated patterns of gaze and gesture were associated with higher cognitive effort and the ongoing development of multimodal control.

Overall, the results highlight the functional role of emotional actions, particularly gestures and gaze, in meaning formation, attention regulation, and listener engagement, confirming that effective interpreting extends beyond linguistic competence to encompass integrated multimodal expertise.

This study makes several important contributions to the emerging field of multimodal translation and interpreting studies.

First, it reveals the dynamic interplay between bodily behaviour and cognitive processes during real-time interpreting, demonstrating how gaze, gestures, and prosody are coordinated to manage attention, regulate cognitive load, and maintain accuracy under stress.

Second, it integrates eye-tracking and gesture analysis within a single multimodal framework, offering a more comprehensive and empirically grounded understanding of interpreter behaviour across perceptual, kinetic, and verbal domains.

Third, it provides empirical support for cognitive-pragmatic and embodied cognitive models by demonstrating that multimodal strategies, particularly the temporal synchronization of gaze, gesture, and speech, enhance both efficiency and communicative accuracy in both simultaneous and consecutive interpreting.

Taken together, these findings support a theoretical basis for viewing interpreting as a bodily, multimodally integrated cognitive activity rather than a purely linguistic task.

Future research should examine how language-specific syntactic structures, information packaging patterns, and cultural traditions of gesture and gaze influence multimodal strategies in diverse interpreting contexts. For instance, interpreters working with languages characterized by flexible word order or high context sensitivity, such as Ukrainian or Japanese, may exhibit different gaze distribution and gesture synchronization compared to those working with languages with fixed syntax and clear grammatical marking, such as English or German.

Such comparative studies would deepen our understanding of how linguistic typology and cultural embodiment interact with the cognitive and pragmatic demands of interpreting. Ultimately, these studies could contribute to the development of a detailed, cross-culturally sensitive model of multimodal interpreting effectiveness, expanding both theoretical and pedagogical frameworks beyond linguistic limitations.

Ethical Considerations

The study was conducted in accordance with general ethical standards of research involving human participants as outlined in the Declaration of Helsinki. All participants gave verbal informed consent after receiving information about the purpose of the study. The data were collected and processed anonymously, ensuring that participants could not be identified. The information obtained was used exclusively for scientific analysis.

References

Cohn, N., Schilperoord, J. (2024). *A Multimodal Language Faculty: A Cognitive Framework for Communication*. London: Bloomsbury. DOI: https://doi.org/10.5040/9781350404861

Boiko Ya. V. (2025). Interpreting in Business: Challenges and Solutions. *Folium,* 6, 32-38, DOI: https://doi.org/10.32782/folium/2025.6.4

Feyaerts, K., Brône, G., Oben, B. (2017). Multimodality in interaction. In B. Dancygier (Ed.), *The Cambridge Handbook of Cognitive Linguistics* (pp. 135-156). Cambridge University Press, DOI: https://doi.org/10.1017/9781316339732.010

Gallagher, S. (2011). *Interpretations of embodied cognition*. In W. Tschacher & C. Bergomi (Eds.), *The Implications of Embodiment: Cognition and Communication (pp. 59-74)*. Exeter: Imprint Academic.

Giberga, A., Ahufinger, N., Igualada, A., Aguilera, M., Guerra, E., Esteve-Gibert, N. (2024). Prosody and gesture in the comprehension of pragmatic meanings: The case of children with developmental language disorder. In Yi. Chen, A. Chen, A. Arvaniti (Eds.), Proceedings of the 12th International Conference on Speech Prosody (pp. 697-701). Leiden: Leiden University Publ. DOI: https://doi.org/10.21437/SpeechProsody.2024-141

Gile, D. (2021). The effort models of interpreting as a didactic construct. In R. Muñoz Martín, S. Sun, D. Li (Eds.), *Advances in Cognitive Translation Studies* (pp. 139-160). Singapore: Springer. DOI: https://doi.org/10.1007/978-981-16-2070-6_7

Hu, T., Wang, X., Xu, H. (2022). Eye-tracking in interpreting studies: A review of four decades of empirical studies. *Frontiers in Psychology*, 13: 872247. DOI: https://doi.org/10.3389/fpsyg.2022.872247

Jelec, A. (2020). Multimodal patterns in cognition and communication. *Studia Anglica Posnaniensia, 55* (s1), 179-184, DOI: https://doi.org/10.2478/stap-2020-0007

Makaruk, L. (2025). Multimodal Syntactic Constructions: A Striking Feature of Digital Communication in Modern English. *Alfred Nobel University Journal of Philology*, 1 (29), 265-283, DOI: https://doi.org/10.32342/3041-217X-2025-1-29-16

Macrine, S., Fugate, J. (2020). *Embodied Cognition*. In K. Hytten (Ed.), *Oxford Research Encyclopedia of Education.* New York: Oxford Academic. DOI: https://doi.org/10.1093/acrefore/9780190264093.013.885

Milošević, J., Risku, H. (2024). Interpreting and embodied cognition. In C.D. Mellinger (Ed.), *The Routledge Handbook of Interpreting and Cognition (pp. 324-340)*. London: Routledge. DOI: https://doi.org/10.4324/9780429297533-24

Oben, B., Brône, G. (2015). What you see is what you do: On the relationship between gaze and gesture in multimodal alignment. *Language and Cognition, 7* (4), 546-562, DOI: https://doi.org/10.1017/langcog.2015.22

Özyürek, A. (2021). Considering the nature of multimodal language from a crosslinguistic perspective. *Journal of Cognition, 4* (1): 42, 1-5. DOI: https://doi.org/10.5334/joc.165

Pinar, M.J. (2013). Multimodality and cognitive linguistics: Introduction to the special volume. *Review of Cognitive Linguistics, 11* (2), 227-235. DOI: https://doi.org/10.1075/rcl.11.2.01pin

Singer, M.A., Radinsky, J., Goldman, S.R. (2008). The role of gesture in meaning construction. *Discourse Processes, 45* (4), 365-386, DOI: https://doi.org/10.1080/01638530802145601

Tiselius, E., Sneed, K. (2020). Gaze and eye movement in dialogue interpreting: An eye-tracking study. *Bilingualism: Language and Cognition, 23* (4), 780-787, DOI: https://doi.org/10.1017/S1366728920000309

Vranjes, J., Brône, G. (2020). Eye-tracking in interpreter-mediated talk: From research to practice. H. Salaets & G. Brône (Eds.), *Linking up With Video: Perspectives on Interpreting Practice and Research* (pp. 203-233). London: Benjamins Translation Library. DOI: https://doi.org/10.1075/btl.149.09vra

Multimodal Integration of Cognitive and Embodied Processes in Interpreting

Oleksandr Sodel, National University of Life and Environmental Sciences of Ukraine (Ukraine)
e-mail: oleksandr_sodel@nubip.edu.ua
DOI: https://doi.org/10.32342/anuJPh.2026.31.23
Key words: *multimodality, interpreting efficiency, gaze, gestures, eye contact, cognitive load, embodied cognition, cognitive processing, multimodal model of interpreting process.*

This study reveals the multimodal integration of cognitive and embodied processes in interpreting by examining the interaction of verbal, visual, and kinetic modalities, focusing on gestures, gaze, and eye contact in meaning formation, managing cognitive load, and communication effectiveness in both simultaneous and consecutive interpreting. The study reviews theoretical framework and previous research on multimodality, analyzes interpreter performance using integrated eye-tracking and gesture analysis, examines the coordination of visual attention and bodily movements at different stages of interpreting, identifies patterns of multimodal synchronization and their cognitive-pragmatic consequences, and proposes a multimodal model of interpreting process that integrates cognitive processing with bodily interaction. Using a mixed-*methods* design, the study analyzed conference speeches, simulated dialogues, and short texts, coding gaze and gestures to examine their coherence with speech and linking visual-kinetic behaviour to performance metrics groups of experienced professional interpreters and graduate student interpreters with different experience levels.

The results reveal a dynamic interplay of bodily actions and cognitive processes. Gestures served as external memory aids: subtle pointing or rhythmic movements facilitated the structuring of information and the memorization of terminology, reducing cognitive load and freeing working memory for comprehension and retrieval. Gaze patterns reflected attentional strategies: experienced professional interpreters proactively anticipated key information, while graduate student interpreters demonstrated a reactive gaze, sometimes delaying paraphrasing. Brief gaze directed at the audience during the transmission indicated monitoring comprehension, maintaining alignment and pragmatic coherence.

Gestures and gaze played a central role in regulating clarity, emphasis, and audience engagement. Rhythmic gestures enhanced prosody and highlighted key terms, synchronized gaze movements marked discourse segmentation, and deictic and iconic gestures facilitated referent tracking and discourse organization. Graduate student interpreters sometimes relied on spontaneous gestures misaligned with the rhythm of speech, which could reduce coherence.

Differences related to mode were identified: simultaneous interpreting required minimalist, rhythmic gestures and brief, strategic gaze fixations due to its high cognitive load, whereas consecutive interpreting allowed for more focused multimodal orchestration, broader gaze movements, and a variety of gestures to structure discourse and enhance expressiveness.

The study presents a multimodal model of interpreting as an integrated system of cognitive and bodily processes, where gaze cues attention, gestures serve as memory aids and communicative tools, and verbal output with prosody interact with these modalities to enhance clarity and coherence. Temporal synchronization of verbal, visual, and kinetic channels ensures effective coordination of multimodal resources, supporting accuracy, fluency, and pragmatic coherence.

Overall, the findings confirm that interpreting is inherently multimodal process, and bodily behaviours serve as integrated cognitive strategies. Effective coordination of modalities improves cognitive control, clarity of delivery, and audience engagement, highlighting the importance of incorporating multimodal strategies into interpreter training.

Дата надходження до редакції / Submitted: 22.12.2025
Дата прийняття до публікації / Accepted: 15.04.2026
Дата публікації / Published: 04.06.2026