Movement Coherence in High Visual Load Environments: Implications for Attention in Mixed-Hearing Classes

Mert Inan[1], Saki Imai[1], Anna Marshall[2], Tessa Karel[2], Malihe Alikhani[1]
[1]Northeastern University [2]Open Up Pittsburgh
{inan.m, m.alikhani}@northeastern.edu

Abstract

Signed interpretation in movement based instruction creates high visual load environments in which spoken language, sign language, and physical demonstration compete for the same perceptual channel. We present a participatory multimodal observational study of mixed hearing movement and mindfulness classes in which Deaf, Hard of Hearing (DHH), and hearing participants practice together. Based on synchronized video recordings, instructor interviews, and participant surveys, we examine how alignment across demonstration, signed instruction, and bodily execution is achieved and restored in real time. Drawing on theories of grounding, repair, and sign language interaction, we conceptualize movement coherence as alignment across these parallel streams and describe how breakdowns trigger observable attention shifts and distributed repair across participants, interpreters, and instructors. Across sessions, we identify recurrent coordination strategies including peer checking, freeze and scan, interpreter repositioning, tactile cueing, and pacing adjustment. Our findings provide an empirically grounded account of grounding under attentional constraint in inclusive embodied settings, with implications for sign language interpretation, multimodal discourse, and the design of accessible movement instruction. This paper includes deidentified materials derived from recorded sessions, including selected keyframes, structured interactional annotations, and anonymized instructor and participant survey responses.

Keywords: movement coherence, Deaf community, participatory design, visual attention, sign language, multimodal communication, repair

1. Introduction

Participation in movement and mindfulness classes requires continuous inference about instructional intent across multiple modalities. What movement does the instructor intend? Has the interpreter accurately conveyed it? Are my neighbors doing what I think I should be doing? In mixed-hearing classes where Deaf, Hard-of-Hearing (DHH), and hearing participants practice together, understanding requires coordinates spoken instruction, signed interpretation and physical demonstration, through a single visual channel. These environments constitute *high visual load* settings, where multiple instruction streams compete for access to the same perceptual channel (Figure 1).

While such environments present challenges for accessibility and instruction, they also provide a unique opportunity to study how people maintain shared understanding when communication unfolds across parallel modalities. In these settings, understanding is not reflected in verbal agreement, but in whether participants' bodies align with the intended movement. When alignment breaks, participants visibly pause, scan the room, reposition themselves, or seek clarification. Interpreters adjust their position or correct movements, and instructors slow down or repeat demonstrations. Alignment is therefore continuously negotiated through observable repair behaviors.

We frame this coordination problem through

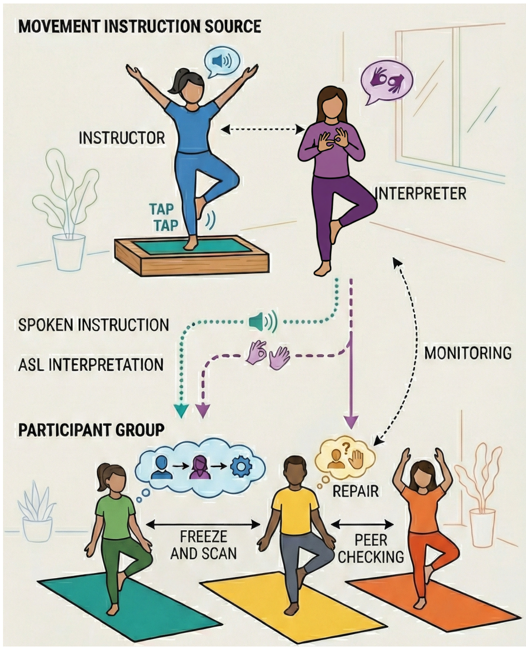

Figure 1: High visual load in mixed-hearing movement instruction. Spoken instruction, signed interpretation, and physical demonstration unfold in parallel, requiring participants to allocate visual attention across competing streams. We model coherence as the alignment between these streams and bodily execution; breakdowns produce observable attention shifts and repair behaviors that are captured in our annotation framework.

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 226–238
16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

Clark's theory of grounding and common ground in communication, which treats mutual understanding as a collaborative and distributed achievement rather than an individual process (Clark, 1996; Clark and Brennan, 1991). In Clark's account, interlocutors continuously coordinate to establish sufficient mutual understanding for the task at hand, initiating repair when alignment falters (Schegloff et al., 1977). In mixed-hearing movement instruction, grounding is not only linguistic but visibly embodied: alignment becomes observable in the coordination of bodies, gaze, and timing across participants, interpreters, and instructors.

Using this theoretical lens, we present a participatory observational study of nine mixed-hearing movement and mindfulness sessions. The study was conducted in collaboration with an inclusive movement and mindfulness education center whose accumulated pedagogical knowledge shaped our observational focus. Using synchronized video recordings, instructor interviews, and participant surveys, we analyze how participants, interpreters, and instructors coordinate attention and repair communication breakdowns in real time. This study is guided by three research questions. First, what coordination challenges arise in mixed-hearing movement instruction under high visual load? Second, how do participants, interpreters, and instructors maintain and repair movement coherence across parallel instruction streams? Third, how are these coordination and repair strategies experienced by participants and instructors?

Our findings show that high visual load environments create a recurring coherence–attention–repair cycle. When instruction streams diverge, participants shift attention, initiate repair through available modalities, and gradually restore alignment. We present a taxonomy of repair strategies distributed across actors and show convergent evidence from video annotation, instructor interviews, and participant surveys. We further identify a tension between comprehension and relaxation in mindfulness contexts, where sustained visual monitoring conflicts with inward attention, and we derive design implications for accessible movement instruction and multimodal communication systems.

This paper makes the following contributions:

1. We identify communication and coordination challenges in mixed-hearing movement instruction through a participatory study with instructors (§ 4).
2. We introduce movement coherence as alignment across demonstration, signed interpretation, and bodily execution, and operationalize it through a video annotation framework that tracks coherence breaks and repair sequences (§ 5).
3. We present a taxonomy of repair strategies initiated by participants, interpreters, and instructors, grounded in the theory of mind inferences each actor makes about others' states (§ 5.2).
4. We connect observational findings with participant and instructor surveys to understand how these repair strategies are perceived and experienced in practice (§ 5.4).

We release deidentified materials derived from recorded sessions, interactional annotations, and anonymized instructor and participant survey responses[1].

2. Related Work

Our account of movement coherence draws on three research traditions: grounding and common ground in communication, coherence relations in multimodal discourse, and visual attention in signed interaction. We synthesize these perspectives to characterize how participants construct coherent understanding across multiple information streams and how they adapt when coherence breaks.

Grounding and Common Ground Understanding in interaction is a collaborative achievement requiring coordination on both content and process (Clark, 1996). Clark and Brennan (1991) analyze grounding as the collective process by which participants reach mutual belief that they have understood one another to a criterion sufficient for current purposes; this account builds on earlier work establishing how speakers and addressees collaborate on contributions (Clark and Schaefer, 1989; Clark and Wilkes-Gibbs, 1986) and how they coordinate through evidence of continued attention (Clark et al., 1983; Schober and Clark, 1989). The effort required for grounding varies with communicative conditions including visibility, audibility, copresence, and shared knowledge (Clark and Brennan, 1991); Isaacs and Clark (1987) show that references between experts and novices require different grounding strategies than those between peers. Enfield (2006) extends this framework by examining the social consequences of common ground. Repair in conversation follows systematic patterns (Schegloff et al., 1977; Schegloff, 1987; Jefferson, 1987), with preference for self-initiated self-repair operating across diverse language communities (Dingemanse et al., 2015). In signed interaction, repair takes modality-specific forms (Manrique and Enfield, 2015; Manrique, 2016): interruption, gestural indication of confusion, and visual confirmation-seeking. Baker (1977) documented how turn-taking in American Sign Language relies on visual attention differently than in spoken discourse.

[1] https://github.com/sakimai/movement-coherence

Coherence Relations in Multimodal Discourse Discourse coherence theory (Hobbs, 1985, 1979) characterizes the inferences giving discourse units coherent joint interpretation through relations such as Elaboration, Explanation, and Parallel; Hobbs and Agar (1985) show how these relations enable coherence even in seemingly discontinuous discourse. Asher and Lascarides (2003) formalize these relations within Segmented Discourse Representation Theory, building on dynamic semantic approaches (Kamp and Reyle, 1993; Groenendijk and Stokhof, 1991). Grosz et al. (1995) model local coherence through relationships among focus of attention, choice of referring expression, and perceived coherence, building on earlier work on focus in dialogue (Grosz, 1977; Sidner, 1983). Centering theory posits that certain entities are more central than others, constraining reference choices (Gordon et al., 1993; Brennan, 1995). Walker et al. (1994) extends centering to Japanese, demonstrating cross-linguistic applicability. Stojnić et al. (2017) argue that pronouns are resolved to whatever entity is at the center of attention in coherent discourse; Stojnić et al. (2020) extend this to demonstratives, showing that pointing gestures update context by putting entities at the center of attention. Coherence relations extend to multimodal communication. Lascarides and Stone (2009) provide a formal semantic analysis showing how speech and gesture jointly contribute through structured relations, building on earlier work on nonverbal communication (Kendon, 2004; McNeill, 1992). Work on image-text coherence (Alikhani et al., 2020; Inan et al., 2021) demonstrates that different relations (Visible, Story, Subjective) affect comprehension and evaluation; Hiippala et al. (2021) extends this to diagrams. Inan and Alikhani (2024) show that coherence relations predict eye-gaze patterns during both comprehension and production, suggesting coherence shapes attentional processes during interpretation.

Visual Attention in Signed Interaction In sign language, coherence is realized through spatial and temporal organization. Inan et al. (2022) show that prosodic features such as intensification and signing space use carry meaning affecting comprehension; generation models failing to capture these features produce outputs signers find incoherent. Inan et al. (2025) address cross-linguistic coherence in sign-to-sign translation, showing alignment algorithms informed by linguistic structure improve quality. Research on sign language interpretation documents cognitive demands on interpreters (Napier, 2009), but less attention has been paid to demands on recipients. Cognitive load theory (Sweller, 1988; Chandler and Sweller, 1991) attributes learning difficulties to instructional designs exceeding working memory capacity, positioning load as an environmental feature. Alibali and Nathan (2014) show that teachers use gesture to reduce cognitive load, building on work linking gesture to spatial reasoning (Alibali et al., 2000; Goldin-Meadow, 2003). Mindfulness practices have documented benefits for DHH populations (Biegel et al., 2009; Black and Slavich, 2015; Werfel et al., 2025), yet these practices assume attentional conditions that high visual load environments disrupt.

Participatory Research Research with DHH communities raises methodological considerations (Singleton et al., 2014; Kusters et al., 2017). Participatory approaches position community members as partners (Bragg et al., 2019; Mack et al., 2020). Our study was conducted in partnership with an inclusive movement and mindfulness education center; community collaborators shaped research questions and interpretive frameworks.

3. Methodology

We conducted a participatory observational study with an inclusive movement and mindfulness education center offering wellness programming for DHH community members. The study combined observations of movement and mindfulness classes, semi-structured instructor interviews, and participant and instructor surveys. These sources allowed us to examine how participants construct understanding of instructor intentions and how communication breakdowns emerge and resolve.

Participatory Partnership We structured the collaboration following principles from Bragg et al. (2019) for research with Deaf communities. Organization staff participated in defining research questions, shaping observation protocols, and interpreting findings. A Deaf community liaison attended all observation sessions. We committed to returning findings in accessible formats and ensuring that any recommendations would serve community-identified needs.

Observational Study We conducted observations following methods for studying embodied interaction in naturalistic settings (Goodwin, 2000; Hutchins, 1995). Observations occurred during regularly scheduled classes over 10 weeks. Each class included eye yoga, gentle movements, and seated mindful breathwork. All classes were led by hearing instructors who provided spoken English instruction with simultaneous ASL interpretation by certified interpreters. Session duration was 50 minutes, with 4 to 10 participants per session. Two synchronized cameras captured complementary perspectives: a front-facing camera documented instructor and interpreter activity while an overview camera captured participants. This dual-view approach enables analysis linking instruction production to participant reception (Mondada, 2006).

Expert & Instructor Interviews To complement observational data, we conducted semi-structured interviews with two instructors and an expert who have led movement and mindfulness programming. Interviews lasted approximately 50 minutes each and addressed teaching challenges, adaptations developed over time, and reflections on what supports or impedes participant engagement. We analyzed interview transcripts thematically, identifying patterns that informed our observational findings.

Participant & Instructor Surveys We additionally collected surveys from both participants and instructors. Participants completed surveys about attention strategies, fatigue, and preferred instructional modalities, while instructors completed post session surveys documenting observed confusion signals, instructional strategies used, and communication challenges encountered during each class. These surveys provided self report evidence that complemented the behavioral patterns observed in the video analysis.

Cohort and Setting Classes operated as mixed-hearing spaces open to Deaf, Hard-of-Hearing, and hearing participants. Our observed cohort comprised 4 participants (age range [18]-[60]; 4 women). Hearing status included Deaf (1), Hard-of-Hearing (1), and hearing (2). Among DHH participants, all used ASL as a primary or secondary communication mode. The mixed-hearing composition created a heterogeneous attention ecology: hearing participants attended primarily to the instructor while DHH participants attended primarily to the interpreter but also monitored the instructor's demonstrations. Different participants thus drew on different information sources for the same instructional content, facing different visual load profiles.

Annotation Protocol We developed the annotation scheme through iterative refinement with community partners and the research team. Each session was segmented into movement episodes corresponding to distinct poses or transitions. Within each episode, we coded four layers: (1) instruction components, marking the onset and offset of spoken description, signed interpretation, and physical demonstration; (2) attention events, including gaze shifts, interpreter repositioning, tactile cues, and eye closure instructions; (3) movement execution states, categorized as aligned, divergent, hesitant, or frozen relative to the demonstrated movement; and (4) repair sequences, marking the initiator (participant, interpreter, or instructor), the trigger (coherence break type), and the outcome (alignment restored or not). The repair categories reported in §5.2 emerged bottom-up from this coding process and were refined over multiple annotation passes.

We used ELAN (Wittenburg et al., 2006) for temporal annotation of the synchronized dual-camera recordings. The first author served as the primary annotator and coded all sessions. To assess reliability, a second annotator independently coded randomly selected movement segments (10 minutes of video) using the same scheme. We computed Cohen's Kappa (Fleiss and Cohen, 1973) for the inter-annotator agreement ($\kappa = 0.67$, *i.e.* substantial agreement). Disagreements were resolved through discussion, and the resulting consensus annotations were used to refine category boundaries for the remaining sessions. The frequency counts reported in Table 1 are based on the primary annotator's full coding of all sessions.

4. Challenges in Mixed-Hearing Movement Contexts

Instructor interviews show how the structural constraints of mixed-hearing environments create distinct cognitive and physiological difficulties for participants that make movement coherence fragile and continuously negotiated. These often place the goals of mindfulness in tension with the visual demands of instruction.

The Paradox of Attention vs. Relaxation A primary difficulty identified by instructors is the conflict between the visual vigilance required for comprehension and the somatic state required for mindfulness. Instructors noted that DHH participants often remained "activated" throughout the entire session. They continuously monitored the visual field for cues rather than "tapping in" to the relaxation experience. Hearing participants might close their eyes or soften their gaze to focus inward. In contrast, DHH participants were observed to physically "strain" to maintain visual contact with the instructor or interpreter. This constant state of alertness prevents the "drop-in" to the body that is central to yoga. The need to externally monitor effectively creates a barrier to internal proprioceptive awareness.

The Validation Loop This heightened activation manifests in what we term a "validation loop." Instructors reported that participants rarely moved on a single cue. Instead, they often required "two to three steps of extra validation" before committing to a pose. A participant might first observe the interpreter and then check the instructor. Finally, they might scan neighboring peers to confirm accuracy. This recursive checking process introduces latency and cognitive load. It disrupts the flow of movement. Instructors attempted to mitigate this by moving around the room to remain visible. Yet the fundamental need for multi-source confirmation remained a persistent tax on participant attention.

Heterogeneity and Pacing Finally, the mixed-hearing cohort presents conflicting demands regarding pacing. Instructors noted a tension between participants who requested a faster flow and those requiring a gentler pace. In a signed environment, speed is a function of visibility. When the pace accelerates, the window for visual processing shrinks. Instructors reported that when the flow became too rapid, the "validation loop" broke down entirely. This caused participants to stop and signal confusion. This suggests that the speed of instruction is limited by the bandwidth of the visual communication channel.

These challenges suggest that coordination in mixed-hearing movement instruction is fundamentally constrained by visual attention and visual bandwidth. Participants must divide attention across multiple information sources while executing movements and engaging in mindfulness practices, making alignment across instruction streams fragile and continuously monitored. When this alignment fails, participants, interpreters, and instructors initiate visible repair actions to restore coordination.

5. Our Framework

The challenges reported by instructors describe the constraints of mixed-hearing movement instruction, where multiple instruction streams compete for visual attention and coordination must be maintained under high visual load. To understand how participants nonetheless coordinate their movements in this environment, we now turn to the movement recordings and apply two concepts to what we observed: *movement coherence* and *theory of mind*. We define both below, then apply them to specific cases from the recordings.

5.1. Embodied Meaning-Making

In the classes we observed, three streams of information unfold in parallel: the instructor's physical demonstration, the interpreter's signed rendering of the spoken instruction, and the participant's own bodily execution.

Definition 1. *We use the term **movement coherence** to refer to the alignment across these three streams: demonstration, signed interpretation, and participant execution. When the instructor extends their arm, the interpreter signs the corresponding description, and the participant extends their arm in the same way, coherence holds. When any of these streams diverge from the others, coherence breaks.*

Algorithm 1 Movement Coherence Parsing

Require: Movement streams for instructor I, interpreter S, participant P; lag threshold τ
Ensure: Coherence state sequence L

Extract movement features M_I, M_S, M_P
for each movement episode t **do**
    Compute alignment and temporal lag between streams
    if streams are aligned and lag $\leq \tau$ **then**
        $L_t \leftarrow$ ALIGNED
    else if streams diverge or lag $> \tau$ **then**
        $L_t \leftarrow$ DIVERGENT
    end if
    if previous state was DIVERGENT and alignment is re-established **then**
        $L_t \leftarrow$ REPAIRING
    end if
end for
return L

To make this construct observable, we annotate coherence at the level of movement episodes using four states: aligned, divergent and repairing. A coherence break is marked when these streams fall out of alignment beyond a brief temporal lag, and a repairing is marked when alignment is re-established. This formulation enables us to track the temporal dynamics of coordination, quantify breakdowns, and relate them to attention shifts and repair strategies across actors. We describe these steps in Algorithm 1. Figure 2(a) shows coherence holding (yellow) across the three actors during aligned segments and breaking (red) when the streams fall out of sync.

Coherence across parallel information streams has been studied in other multimodal domains. Cumming et al. (2017) describe conventions that maintain spatial coherence across discontinuous film shots; Lascarides and Stone (2009) formalize coherence between gesture and speech; Alikhani et al. (2020) model coherence between images and their textual descriptions. We extend this line of work to a triadic embodied setting where demonstrations, signs, and bodily movements must cohere across three actors.

Coherence is also tied to attention. Stojnić et al. (2017) show that in linguistic discourse, coherence relations govern what is at the center of attention at any given point, and that attention determines how context-sensitive expressions are resolved. We observe the same structure in our data. When coherence holds, participants distribute attention across streams. When it breaks, attention narrows to a single stream that can resolve the mismatch. Figure 2(b) shows this cycle: coherence breaks trigger attention shifts, which trigger repair, which restores coherence.

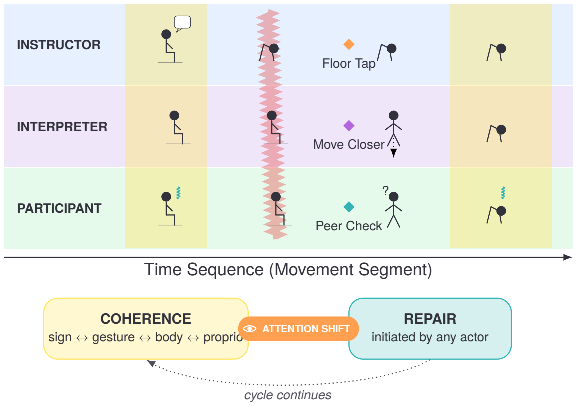

Figure 2: (a) Coherence (yellow) holds when signed instruction, demonstration, and bodily execution align across actors. Breaks (red) trigger attention shifts and repair actions (diamonds). (b) The coherence-repair cycle: attention shift detects breaks and initiates repair by any actor.

Coherence describes whether the three streams are aligned. But it does not explain how people figure out what to do when they come apart. For that, we need theory of mind (ToM): the capacity to infer what others intend and believe (Premack and Woodruff, 1978; Wimmer and Perner, 1983; Apperly, 2010; Schaafsma et al., 2015). Each actor in the room continuously models the others' intentions. The participant infers what movement the instructor wants them to do and whether the interpreter has accurately conveyed it. The interpreter monitors the participant's execution and infers their comprehension state. The instructor reads the group's movements to gauge whether the class is following. When coherence breaks, these inferences drive repair (Clark, 1996; Clark and Brennan, 1991; Schegloff et al., 1977). A participant who points to their own leg is signaling confusion; the interpreter who points back has inferred what the participant needs to know; an instructor who pauses has inferred that the group has fallen behind.

Clark's account of grounding provides a framework for understanding how participants collaboratively establish mutual understanding through visible coordination and repair. However, it remains largely interactional and does not explicitly model how individuals infer others' intentions when coordination fails. We therefore complement grounding with ToM, which captures the inferential processes by which participants interpret others' beliefs and intentions during breakdowns. In our setting, grounding is visible in the alignment of movements across actors, while theory of mind explains how participants decide what corrective action to take when this alignment is disrupted.

We observed this coherence-attention-repair sequence across multiple instances. However, interviews with instructors and experts suggest that these repair sequences are not occasional events but structural features of mixed-hearing movement instruction. Because multiple instruction streams compete for the same visual channel, maintaining coherence requires continuous monitoring, and coherence is frequently placed under pressure.

5.2. Repair Strategies

Through a bottom-up, data-driven approach we present a taxonomy of repair strategies (see Table 1). The categories emerged from iterative annotation of the video data and are not an exhaustive list. We organized them by who initiates the repair: participant, interpreter, or instructor.

Figure 3 illustrates a scenario where several of these strategies co-occur. During an eye yoga exercise, participants were instructed to focus on their thumb while shifting visual focus between the thumb and a point behind it. The exercise required sustained visual attention on a specific target while instruction about when to shift focus arrived through the same visual channel. Participants adapted by checking the interpreter between focus shifts, relying on peripheral vision, or following peer timing.

Participant-initiated repairs. When visual load exceeded capacity, participants employed adaptive strategies to restore alignment. Some are self-initiated repairs in the conversation-analytic sense (Schegloff et al., 1977); others are clarification-seeking actions directed at another actor (Clark, 1996). Both are triggered by coherence breaks and aim to restore alignment.

- **Peer-checking**: observing neighbors as a proxy for correct movement (Goodwin, 2000).
- **Freeze-and-scan**: pausing movement to sweep gaze across information sources (Kendon, 1967).
- **Strategic disengagement**: reducing monitoring during familiar sequences to conserve attentional resources (Hutchins, 1995).
- **Position-holding**: maintaining a previous pose during uncertainty (Clark, 1996).
- **Signing to the interpreter**: producing a gesture or sign directed at the interpreter to request clarification, such as pointing to a body part to ask “this one?” (Schegloff et al., 1977).
- **Communicating with peers or instructor**: verbally or gesturally addressing another actor for confirmation or request a repeat, showing uncertainty (Clark and Wilkes-Gibbs, 1986).

Interpreter-initiated repairs. Interpreters actively monitored participant behavior and adjusted their own positioning and output to restore alignment. These are other-initiated repairs (Schegloff et al., 1977) that operate across the visual and spatial modalities available in the signed environment.

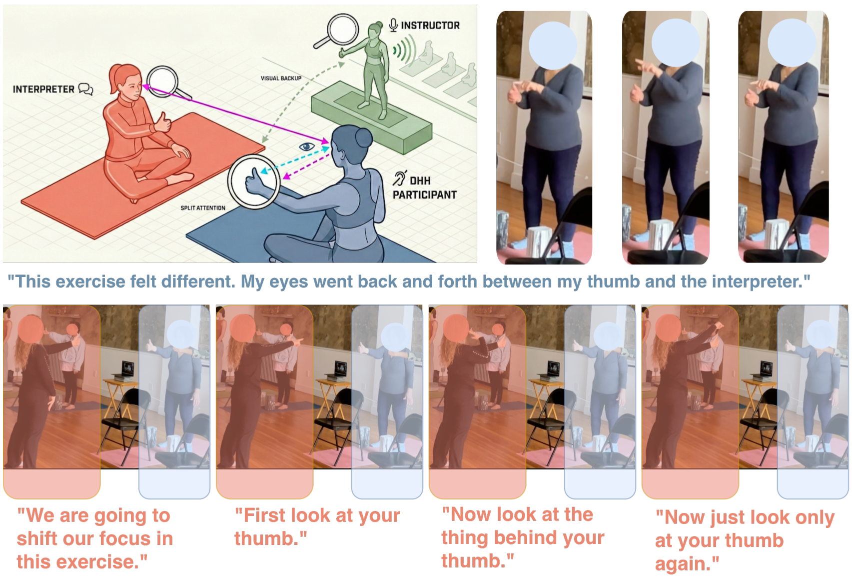

Figure 3: The eye yoga exercise illustrates visual load exceeding capacity at its most acute. The practice demands sustained visual focus on one's thumb while instruction about when to shift focus arrives through the same visual channel. Participants adapt by checking the interpreter between focus shifts, relying on peripheral vision, or following peer timing.

Repair Strategy	**Initiator**	**Trigger**	**Modality**	**N**
Peer-checking	Participant	Uncertainty about correct movement	Visual (gaze to neighbors)	7
Freeze-and-scan	Participant	Maximum ambiguity at transition	Visual (gaze sweep)	3
Strategic disengagement	Participant	Familiar or low-stakes sequence	Visual (gaze softening)	5
Position-holding	Participant	Uncertain transition	Bodily (maintain prior pose)	2
Signing to interpreter	Participant	Need for explicit clarification	Gestural/signed (deictic, question)	3
Communicating with peers/instructor	Participant	Uncertainty, need for confirmation	Verbal/gestural (speech, sign, gesture)	2
Proximity adjustment	Interpreter	Participant posture limits view	Spatial (repositioning)	11
Persistent signing	Interpreter	Temporarily outside participant view	Linguistic (continued signing)	11
Error correction	Interpreter	Participant executes wrong movement	Gestural (pointing, demo)	4
Strategic non-intervention	Interpreter	Participant visually occluded	None (waiting)	4
Tactile cueing	Instructor	Participant cannot see visual cues	Tactile (floor tap)	4
Dual monitoring	Instructor	Ongoing instruction	Visual (checking interp. + participants)	10
Pacing adjustment	Instructor	Participant appears confused	Temporal (slowing, repeating)	4

Table 1: Taxonomy of repair strategies observed in mixed-hearing movement classes, organized by initiator. Frequency counts (N) are from ELAN annotations across all sessions. Repair is distributed across all three actors, with each actor drawing on the modalities available to them.

- **Proximity adjustment**: moving into the participant's restricted visual field when their posture limits their view (Metzger, 1999).
- **Persistent signing**: continuing to sign even when temporarily outside the participant's view, maintaining the linguistic stream for when visual access resumes (Winston, 1994).
- **Error correction**: noticing a participant's execution error (e.g. wrong side in a mirrored movement) and producing a gestural correction (Schegloff et al., 1977).
- **Strategic non-intervention**: waiting for a participant to complete a visually occluded movement, recognizing that interruption may be more disruptive than partial information (Napier, 2002).

Instructor-initiated repairs. Instructors monitored both the interpreter and participants for signs of misalignment and adjusted the pace or modality of instruction accordingly.

- **Tactile cueing**: tapping on the floor to signal transitions through a non-visual channel, enabling communication regardless of visual orientation (Jain et al., 2015).
- **Dual monitoring**: checking both the interpreter's rendering and the participants' execution to verify that information is reaching the group (Clark, 1996).
- **Pacing adjustment**: slowing instruction or repeating a demonstration when participants appear confused, providing additional processing time (Schegloff et al., 1977).

We also observed *error propagation*: the interpreter accidentally performed a movement the instructor had not demonstrated, and the participant copied it. Coherence between participant and interpreter held, but coherence with the instructor's actual instruction was broken. Repair mechanisms can introduce new coherence breaks when the participant cannot verify against multiple sources.

The repair strategies documented above all assume that visual access, even if limited, remains partially available. Participants can still glance at the interpreter, check a neighbor, or scan the instructor. The next section examines what happens when this assumption breaks down entirely.

5.3. Eyes-Closed: Visual Constraint at Maximum

Mindfulness components that instructed participants to close their eyes represented the most extreme form of visual constraint. With eyes closed, DHH participants lost access to their primary communication channel entirely. We observed 15 eyes-closed segments in which actors developed coordinated strategies to maintain coherence.

Before eyes-closed segments, participants caught the interpreter's eye and established which signs would indicate breathing in versus breathing out (*pre-occlusion coordination*). During forward folds and other positions that occluded the participant's view, the interpreter moved into the participant's restricted visual field (*interpreter proximity adjustment*). During fully occluded segments, the instructor tapped the floor to provide temporal structure (*tactile timing*). Participants varied in their response: some closed eyes fully, trusting the pre-established coordination; others kept eyes partially open, prioritizing comprehension monitoring over full participation in the mindfulness component.

5.4. Evidence from Instructor and Participant Surveys

While our observational analysis characterizes how coherence is maintained and repaired in real time, surveys from participants and instructors provide insight into how these dynamics are experienced and interpreted in practice. These reports reveal which aspects of the coherence–attention–repair cycle are perceptually salient to different actors and how they shape participation in the class.

Instructor observed confusion signals algin with repair behaviors. Instructor reports of confusion signals align closely with the repair strategies identified through video annotation. Instructors reported observing scattered gaze (4 sessions), peer copying (3 sessions), and uncertain posture or fidgeting (3 sessions) which are behavioral signals that correspond precisely to freeze-and-scan, peer-checking, and position-holding in Table 1 (see Figure 4, rows 1-3). This convergence strengthens the

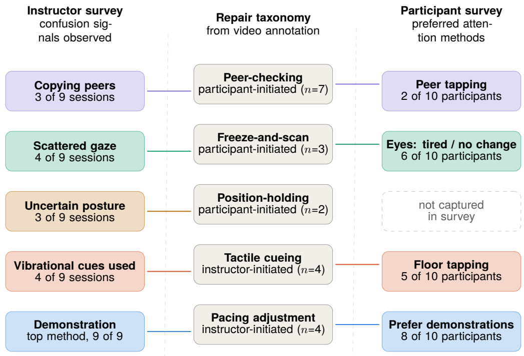

Figure 4: Convergence across instructor surveys, video annotated repair taxonomy, and participant surveys. Each row maps an instructor observed confusion signal to the corresponding repair strategy in Table 1 and, where captured, to the participant-endorsed attention method.

validity of the repair taxonomy and suggests that instructors recognize the same coherence breakdown signals that are visible in the recordings.

Instructional strategies that reduce visual monitoring. Surveys showed that both instructors and participants preferred instructional methods that reduced reliance on signed or verbal explanation, such as physical demonstration, floor tapping, and vibrational cues (Figure 4, rows 4-5). While demonstration is still visual, it differs from signed instruction in that participants can follow movement directly through embodied imitation rather than sustained linguistic attention. Floor tapping and vibrational cues further distribute information into non-visual channels. These strategies reduce the need for continuous monitoring of linguistic instruction.

Fatigue reflects sustained coherence monitoring. Post-class survey responses regarding eye fatigue further support the claim that participants must maintain sustained visual monitoring to preserve coherence. Participants reported varying levels of eye fatigue, suggesting heterogeneity in how much visual monitoring individuals maintain during class. This supports the interview findings that some participants prioritize comprehension monitoring while others prioritize relaxation, which reinforces the tension between attention and mindfulness identified earlier.

6. Discussion and Implications for Future Design

Movement coherence extends grounding theory to embodied contexts. Our observations and interviews extend grounding theory (Clark, 1996) to embodied, high visual load contexts. In

Clark's framework, successful communication requires continuous coordination between interlocutors to establish mutual understanding. Our data show how this coordination operates when linguistic and physical information streams compete for the same attentional channel, and when the goal of communication is bodily alignment rather than propositional agreement. Movement coherence provides a visible, continuous signal of grounding success or failure, and the repair sequences we documented show how interlocutors use these signals to detect and resolve breakdowns. Repair in this setting differs from spoken conversation, where it typically addresses a specific trouble source in the linguistic stream. In movement instruction, repair addresses misalignment between multiple parallel streams (demonstration, interpretation, execution), and any interlocutor can initiate it based on the signals available to them.

Visual reliance creates a specific burden in mindfulness contexts. Because DHH individuals communicate through sustained visual attention, their eyes are under constant cognitive and physical strain during instruction. Survey responses and instructor observations both indicate that participants experience visual fatigue and must continuously monitor multiple visual streams to maintain alignment. Mindfulness instruction asks participants to soften external monitoring and attend inward, while maintaining movement coherence requires sustained monitoring.

Stillness and rest function as repair strategies. Expert and instructor interviews suggest that moments of stillness function as a repair strategy for the nervous system. When visual and cognitive demands accumulate, participants need structured breaks to reset. Our expert recommended 3 to 10 minute breaks of stillness with eyes closed, focused on body awareness, after which participants perform differently. If comprehension monitoring and relaxation are incompatible within continuous instruction, then alternating between active instruction and structured rest may be more effective than attempting both simultaneously.

Design implications. Our findings suggest several implications for the design of inclusive movement instruction in high visual load environments:

- **Consistent structure reduces comprehension monitoring.** Repeated class structure allows participants to anticipate coherent movement and reduce visual monitoring, redirecting cognitive resources toward bodily experience.
- **Instruction should distribute information across sensory modalities and representations.** Our expert recommended activating tactile channels (hand-holding for connection, tapping chains for transitions, self-touch during breathwork), vibrational channels (gongs or vibrating instruments, smartwatch vibrations for timing cues), and proprioceptive engagement (grounding through seated positions, balancing poses). These methods reduce dependence on signed linguistic description and instead use embodied demonstration that participants can follow without sustained visual fixation on an interpreter.
- **Peer pairing distributes the work of maintaining coherence.** Pairs allow participants to observe and correct each other at close range, suggesting that dyadic models (e.g., an avatar guiding two paired participants) may be more effective than broadcast instruction.
- **Technology could support personalization and instructor awareness.** The behavioral markers we identified, such as freezing, scanning, peer-checking, and movement divergence, could be used by multimodal systems to detect coherence breakdowns in real time. Technologies such as AR overlays, avatars (Stoll et al., 2020), or haptic feedback devices could help distribute timing and instruction across visual and non-visual channels, allowing participants to maintain synchronization without continuous visual monitoring.

7. Conclusion

By observing movement coherence in mixed-hearing instruction, we provide an analysis framework grounded in everyday embodied communication. We consistently observe that when demonstration, interpretation, and participants' movements fall out of sync, someone in the room initiates a visible repair: participants pause and scan, interpreters reposition or correct, instructors slow down or repeat. Alignment is actively rebuilt in real time. We also observe a tension between comprehension and relaxation, where sustained visual monitoring limits the inward attention that mindfulness practices are designed to support. Expert and instructor interviews point toward concrete design principles for addressing these challenges: an asset-based orientation, multi-modal instruction, consistent class structure, peer pairing, and structured stillness. Throughout this work, participatory design has been essential to identifying challenges that matter to the DHH community and will remain essential to developing solutions that serve community needs. With the framework and data released through this study, we hope to spearhead research into multimodal communication specifically for *in situ* movement and mindfulness settings in mixed-hearing environments.

8. Limitations

Our cohort was drawn from a single organization, limiting generalizability across DHH communities, regional signing varieties, and class formats. Deaf-led classes without interpretation would likely show different dynamics. Our analysis privileges observable behavior; participants' subjective experience may differ from our interpretations. The participatory framework shaped research questions toward community priorities, potentially limiting attention to phenomena of primarily academic interest.

9. Ethical Considerations

This study was conducted under IRB approval (#25-08-35). All participants provided informed consent, and all data were deidentified prior to analysis and publication. We took care to minimize any potential risks to participants, and no harm resulted from the study.

10. Acknowledgements

We thank Open Up Pittsburgh and all the participants of the classes who volunteered for us to use their data. We thank Megan Flinn for their guidance and thoughtful advice on the study design, development of research questions, and feedback throughout the project.

11. Bibliographical References

Martha Alibali, Sotaro Kita, and Amanda Young. 2000. Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes - LANG COGNITIVE PROCESS*, 15:593–613.

Martha W Alibali and Mitchell J Nathan. 2014. Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. In *Video research in the learning sciences*, pages 349–365. Routledge.

Malihe Alikhani, Piyush Sharma, Shengjie Li, Radu Soricut, and Matthew Stone. 2020. Cross-modal coherence modeling for caption generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6525–6535, Online. Association for Computational Linguistics.

Ian Apperly. 2010. *Mindreaders: The cognitive basis of "Theory of Mind"*. Psychology Press, Hove, UK.

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge, UK.

Charlotte Baker. 1977. Regulators and turn-taking in American Sign Language discourse. *On the other hand*, pages 215–236.

Gina M Biegel, Kirk Warren Brown, Shauna L Shapiro, and Christine M Schubert. 2009. Mindfulness-based stress reduction for the treatment of adolescent psychiatric outpatients: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 77(5):855–866.

David S Black and George M Slavich. 2015. Mindfulness meditation and the immune system: A systematic review of randomized controlled trials. *Annals of the New York Academy of Sciences*, 1373(1):13–24.

Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign language recognition, generation, and translation: An interdisciplinary perspective. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19, page 16–31, New York, NY, USA. Association for Computing Machinery.

Susan E. Brennan. 1995. Centering attention in discourse. *Language and Cognitive Processes*, 10(2):137–167.

Paul Chandler and John Sweller. 1991. Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4):293–332.

Herbert H. Clark. 1996. *Using Language*. "Using" Linguistic Books. Cambridge University Press.

Herbert H Clark and Susan E Brennan. 1991. Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149.

Herbert H Clark and Edward F Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13(2):259–294.

Herbert H Clark, Robert Schreuder, and Samuel Buttrick. 1983. Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, 22(2):245–258.

Herbert H Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Samuel Cumming, Gabriel Greenberg, and Rory Kelly. 2017. Conventions of viewpoint coherence in film. *Philosophers' Imprint*, 17(1):1–29.

Mark Dingemanse, Seán G Roberts, Julija Baranova, et al. 2015. Universal principles in the repair of communication problems. *PloS one*, 10(9):e0136100.

Nicholas J Enfield. 2006. Social consequences of common ground. In *Roots of Human Sociality: Culture, Cognition and Interaction*, pages 399–430. Berg, Oxford.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3):613–619.

Susan Goldin-Meadow. 2003. Hearing gesture: How our hands help us think.

Charles Goodwin. 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10):1489–1522.

Peter C. Gordon, Barbara J. Grosz, and Laura A. Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17(3):311–347.

Jeroen Groenendijk and Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy*, 14(1):39–100.

Barbara J. Grosz. 1977. The representation and use of focus in a system for understanding dialogs. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 1*, IJCAI'77, page 67–76, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Tuomo Hiippala, Malihe Alikhani, Jonas Haverinen, Timo Kalliokoski, Evanfiya Logacheva, Serafina Orekhova, Aino Tuomainen, Matthew Stone, and John A. Bateman. 2021. AI2D-RST: a multimodal corpus of 1000 primary school science diagrams. *Lang. Resour. Eval.*, 55(3):661–688.

Jerry R Hobbs. 1979. Coherence and coreference. *Cognitive Science*, 3(1):67–90.

Jerry R Hobbs. 1985. On the coherence and structure of discourse. In Livia Polanyi, editor, *The Structure of Discourse*. Ablex, Norwood, NJ.

Jerry R Hobbs and Michael H Agar. 1985. The coherence of incoherent discourse. *Journal of Language and Social Psychology*, 4(3-4):213–232.

Edwin Hutchins. 1995. *Cognition in the Wild*. MIT Press, Cambridge, MA.

Mert Inan and Malihe Alikhani. 2024. Seeing eye-to-eye: Cross-modal coherence relations inform eye-gaze patterns during comprehension & production. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 14494–14512, Torino, Italia. ELRA and ICCL.

Mert Inan, Piyush Sharma, Baber Khalid, Radu Soricut, Matthew Stone, and Malihe Alikhani. 2021. COSMic: A coherence-aware generation metric for image descriptions. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3419–3430, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Mert Inan, Yang Zhong, and Malihe Alikhani. 2025. How to align multiple signed language corpora for better sign-to-sign translations? In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4003–4016, Albuquerque, New Mexico. Association for Computational Linguistics.

Mert Inan, Yang Zhong, Sabit Hassan, Lorna Quandt, and Malihe Alikhani. 2022. Modeling intensification for sign language generation: A computational approach. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2897–2911, Dublin, Ireland. Association for Computational Linguistics.

Ellen A Isaacs and Herbert H Clark. 1987. References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1):26–37.

Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. 2015. Head-mounted display visualizations to support sound awareness for the deaf and hard of hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 241–250, New York, NY, USA. Association for Computing Machinery.

Gail Jefferson. 1987. On exposed and embedded correction in conversation. In *Talk and Social Organisation*, pages 86–100. Multilingual Matters.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht.

Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63.

Adam Kendon. 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press.

Annelies Kusters, Maartje De Meulder, and Dai O'Brien. 2017. *Innovations in deaf studies: The role of deaf scholars*. Oxford University Press.

Alex Lascarides and Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics*, 26(4):393–449.

Kelly Mack, Danielle Bragg, Meredith Ringel Morris, Maarten W. Bos, Isabelle Albi, and Andrés Monroy-Hernández. 2020. Social app accessibility for deaf signers. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW2).

Elizabeth Manrique. 2016. Other-initiated repair in Argentine Sign Language. *Open Linguistics*, 2(1).

Elizabeth Manrique and Nicholas J Enfield. 2015. Suspending the next turn as a form of repair initiation: Evidence from Argentine Sign Language. *Frontiers in Psychology*, 6:1326.

David McNeill. 1992. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

Melanie Metzger. 1999. *Sign Language Interpreting: Deconstructing the Myth of Neutrality*. Gallaudet University Press, Washington, DC.

Lorenza Mondada. 2006. Video recording as the reflexive preservation and configuration of phenomenal features for analysis. In *Video analysis: methodology and methods : qualitative audio-visual data analysis in sociology*, pages 51–68. Lang.

Jemina Napier. 2002. *Sign Language Interpreting: Linguistic Coping Strategies*. Douglas McLean, Coleford, UK.

Jemina Napier. 2009. *Exploring linguistic and cultural identity: My personal experience*.

David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4):515–526.

Sara M. Schaafsma, Donald W. Pfaff, Robert P. Spunt, and Ralph Adolphs. 2015. Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, 19(2):65–72.

Emanuel A Schegloff. 1987. Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. *Talk and Social Organisation*, pages 70–85.

Emanuel A Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.

Michael F Schober and Herbert H Clark. 1989. Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2):211–232.

Candace L Sidner. 1983. Focusing in the comprehension of definite anaphora. *Computational Models of Discourse*, pages 267–330.

Jenny L. Singleton, Gabrielle Jones, and Shilpa Hanumantha. 2014. Toward ethical research practice with deaf participants. *Journal of Empirical Research on Human Research Ethics*, 9(3):59–66. PMID: 25746786.

Una Stojnić, Matthew Stone, and Ernest Lepore. 2017. Discourse and logical form: pronouns, attention and coherence. *Linguistics and Philosophy*, 40(5):519–547.

Una Stojnić, Matthew Stone, and Ernest Lepore. 2020. Pointing things out: In defense of attention and coherence. *Linguistics and Philosophy*, 43(2):139–148.

Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, and Richard Bowden. 2020. Text2Sign: Towards sign language production using neural machine translation and generative adversarial networks. In *International Journal of Computer Vision*, volume 128, pages 891–908. Springer.

John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285.

Marilyn A Walker, Masayo Iida, and Sharon Cote. 1994. Japanese discourse and the process of centering. *Computational Linguistics*, 20(2):193–232.

Krystal L. Werfel, Jessica Mattingly, and Emily Lund. 2025. Yoga as a fatigue intervention for children who are deaf and hard of hearing. *Perspectives of the ASHA Special Interest Groups*, 10(4):1180–1189.

Heinz Wimmer and Josef Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1):103–128.

Elizabeth A Winston. 1994. An interpreted education: Inclusion or exclusion? *Gallaudet University Communication Forum*, 3:55–62.

12. Language Resource References

Wittenburg, Peter and Brugman, Hennie and Russel, Albert and Klassmann, Alex and Sloetjes, Han. 2006. *ELAN: A professional framework for multimodality research*. The Language Archive, Max Planck Institute for Psycholinguistics.