UNIVERSITY OF GENOVA

PHD PROGRAM IN BIOENGINEERING AND ROBOTICS

The hidden dimension of connection: the development of intersubjectivity in infants and children with visual impairments

by

Marta Guarischi

Thesis submitted for the degree of *Doctor of Philosophy* (XXXVIII cycle)

June 2026

Monica Gori	Supervisor
Eleonora Montagnani & Giulia Cappagli	Co-supervisor
Paolo Massobrio	Head of the PhD program

Thesis Jury:

Andrew J. Bremner, *University of Birmingham*	External examiner
Dorothy Cowie, *Durham University*	External examiner

Dibris

Department of Informatics, Bioengineering, Robotics and Systems Engineering

I am grateful to all those who have followed the winding path of my work, and to those who patiently waited for its completion. As for those who believe that raising questions, starting over, searching, making mistakes, revisiting everything from top to bottom, and still finding reasons to hesitate at every step: in short, for those to whom working through difficulties and living in constant intellectual struggle appears as a kind of resignation, well, we clearly do not belong to the same planet.

.

Michel Foucault- read, experienced and embodied

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Marta Guarischi

April 2026

Acknowledgements

To my *parents*, who remind me that becoming oneself requires crossing the hidden dimension of connection and to *Paolo*, with whom I explore its depths.
I am deeply grateful to the *children and parents* of the Mondino Foundation, who taught me to look beyond what we are used to seeing. To my supervisor, *Monica*, who from the very beginning showed me that creativity is the essential foundation of research.
To *Eleonora*, with whom collaboration feels like a reciprocal adjustment: a familiarity shaped by knowing each other in three-dimensional space. To *Giulia*, who helped me refine my direction with method, and with that subtle touch of madness she taught me to channel.
I am also grateful to the clinical counterpart, the *Mondino club,* who helped me translate intuitions into other languages they might otherwise never have spoken. To *Giacomo* with whom life is a funny dance for two. To *Giacomo*, with whom life is a funny dance for two.
To the "*Bulle"*, travelling companions without whom I would have lost four luminous angles from which to see the world.

Scientific production

Peer-reviewed journal papers:

- **Guarischi**, M., Montagnani, E., Catalano, G., Saligari, E., Signorini, S., & Gori, M. (2025). From motion to interaction: How multisensory information shapes motor behaviors in children with visual impairment. Research in Developmental Disabilities, 159, 104956. doi: 10.1016/j.ridd.2025.104956
- **Guarischi** M., Balzarotti N., Cappagli G., Campus C., Morelli F., Catalano G., Signorini S., Gori M. (2025). "Beyond race model predictions: The temporal dynamics of multisensory attentional capture", Acta Psychologica, 261, 105833. doi: 10.1016/j.actpsy.2025.105833
- Bertolasi J., Garcia-Hernandez N.,, Memeo M., **Guarischi** M., Gori M. (2025). Evaluation of HoloLens 2 for Hand Tracking and Kinematic Features Assessment. Virtual Worlds 4(3), 31. doi:10.3390/virtualworlds4030031
- Coelho L., Aguilar D., Basta S., **Guarischi** M., Gonzalez C., Gori M. (2025). The role of visual experience in haptic spatial perception: evidence from early blind, late blind, and sighted individuals. Biology of Sex Differences Biology of Sex Differences16(1). doi: 10.1186/s13293-025-00747-y
- Catalano, G., Abbondio, S., Nicotra, R., Berselli, V., **Guarischi**, M., Vezzali, V., & Signorini, S. (2026). Eye Tracking for Rehabilitation and Training in Paediatric Neurodevelopmental Disorders: A Systematic Review. *Brain Sciences*, *16*(3), 337. doi: 10.3390/brainsci16030337
- Montagnani M., Senacheribbe A., Mancinelli C., Beukeboom R., **Guarischi** M., Gori M. "Overcoming boundaries of environmental probing with sp.ex: a parameter-free quantification of spatial exploration" *Under review*
- **Guarischi** M., Montagnani E., Memeo M., Catalano G., Dagnino C., Noceti N., Signorini S., Gori M. "The angles between us: modeling embodied joint attention episodes in mother-child dyads" *Under review*
- **Guarischi** M. & Cappagli G., Catalano G., Signorini S., Gori M. "Visual deprivation alters caregiver-specific autonomic tuning during infancy" *Under review*
- Montagnani E. & **Guarischi** M., Nicotra R., Catalano G., Memeo M., Gori M., Signorini S. "Variability matters: effects of real-time sonification in supporting spatial orientation of children with and without vision" *Under review*

Peer-reviewed conference papers:

- Morelli, F. Balzarotti, N. **Guarischi**, M. Cappagli, G. Maviglia, A. Crepaldi, M. Orciari, L. Parmiggiani, A. Catalano, G. Signorini S. & Gori M.. "A novel multisensory device for the assessment and rehabilitation of perceptual and attentional competencies" in 2023 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, doi: 10.1109/EMBC40787.2023.10340824
- **Guarischi** M. & Hu, Y., Kurt A.B., Zanchi S., Barth E., Gori M. "A machine learning approach to unveil balance behavior through aging with auditory cue", in 2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA), doi: 10.1109/MeMeA60663.2024.10596824
- Montagnani E & Memeo M.., **Guarischi** M., Catalano G., Saligari E., Signorini S., Gori M. "SONIA: A Real-Time Automated Framework for Sonification of Navigation and Interactive Activities for Children with Visual Impairment" in 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) doi: 10.1109/EMBC58623.2025.11251721

Posters:

- Perceiving more than expected: the deviation of responses to bimodal stimuli from the race model *at International Multisensory Research Forum (IMRF), 2023.*
- A novel multisensory device for the assessment and rehabilitation of perceptual and attentional competencies *at Child Vision Research Society (CVRSoc), 2023.*
- "What do you see?": the role of vision in the development of allocentric perspective *at Associazione Italiana di Psicologia (AIP), 2023.*
- A machine learning approach to unveil balance behavior through aging with an auditory cue *at IEEE Medical Measurements & Applications (MeMeA), 2024*
- The ecological spatial exploration: how sense guide the locomotion of children with visual impairments *at European Society for Movement Analysis in Adults and Children (ESMAC), 2024* ***Best Poster Award Nomination***
- Feeling the body: hand representation in early and late blind adults *at Body Representation Network (BRnet), 2025*
- The angles between us: quantifying relational, directional, joint and separate attention *at Joint Action Meeting (JAM), 2025*
- The angles between us: modeling embodied joint attention episodes in caregiver-child dyads *at International Workshop on Naturalistic Experimentation of Child Development (NECD), 2026*

Talks:

- The dance of attachment: linking motor characteristics with attachment and behavioural measures of the dyad mother-child with visual impairment *at Associazione Italiana di Psicologia (AIP), 2024*
- Tracing the vectors of attention, *at Department of Psychology, University of Copenaghen, 2025*
- Kinematic measures for relational outcomes in visually impaired children *at the International Alliance of Academies of Childhood Disability (IAACD) and Annual Meeting of the European Academy of Childhood-onset Disability (EACD), 2025*
- The geometry of connection: mapping head orientation in visually impaired children and their caregivers *at Associazione Italiana per le Scienze Cognitive (AISC), 2025*

Workshop:

- Coordinator of the activity "Non si vede solo con gli occhi", *Festival della Scienza, 2024*
- Gait Course, European Society for Movement Analysis in Adults and Children *(ESMAC), 2024*
- Summer school, Nonlinear Analysis Workshop, *University of Nebraska, 2025*
- International Workshop on Naturalistic Experimentation of Child Development *(NECD), 2026* ***Travel Grant***

Patent:

- Italian Patent Application No. 102025000009990 "Method and system for quantifying the spatial exploration of a region of interest by a subject." Inventors: Montagnani E., Senachetibbe A., Mancinelli C., **Guarischi** M., Gori

Abstract

Intersubjectivity is conceptualized as a dynamic, embodied process that emerges during growth through multisensory engagement and motor development. The present thesis aims to adapt developmental protocols to ensure inclusive models for infants and children with visual impairment (VI). These adaptations replace gaze-dependent measures with indicators derived from alternative sensory channels, such as touch, proprioception and audition. To investigate how visual experience shapes the reorganization of intersubjective dynamics during early development, the project adopted a multisensory and technology-driven approach articulated across four E-dimensions: Embodied, Embedded, Enculturated and Enabling. Embodied intersubjectivity investigates how visual deprivation is associated with differences in the discrimination of affective touch, revealing preliminary evidence of shifts in caregiver-specific physiological tuning in infants with VI. Embedded intersubjectivity adapts a protocol inspired by the Strange Situation paradigm to children with VI and explores how unexpected multisensory input could modulate kinematic variability and proximity to caregivers, suggesting that flexibility in attention might function as a strategy for environmental control when visual information is limited. Enculturated intersubjectivity develops and validates a multimodal framework for detecting episodes of joint attention through head orientation during naturalistic caregiver and child interactions, integrating clinical video coding, computer vision, and marker-based motion capture. Enabling intersubjectivity presents an innovative sonification system that translates movement into sound in real time, suggesting that auditory feedback can support proprioception, scaffolding motor exploration in the absence of vision. Across these four dimensions, the findings suggest that in infants and children with VI autonomic responses might shift, motor variability could play an adaptive, head orientation provide a potential operational index of joint attention, and auditory cues might scaffold navigation and bodily awareness. The significance of this thesis lies in its strong translational potential. It suggests that interactive research paradigms traditionally rooted in vision can be redesigned around residual sensory abilities, and that technological tools can be used to both detect and support engagement and scaffold navigation. Together, these contributions have the potential to open new avenues for research and clinical practice. A crucial application concerns the transition from qualitative to quantitative assessment of relational indicators, leveraging time-series analysis to capture the dynamic nature of interactive processes. Future developments will aim to extend these metrics to other clinical populations, as relational processes constitute a domain of investigation across neurodevelopmental disorders. Intersubjectivity is far from being a static trait and emerges as a dynamic relational system, offering not only a powerful lens to understand development, but a new way of seeing when vision itself is compromised.

Table of contents

List of Figures

List of Tables

Chapter 1

Background

1.1 Intersubjectivity and its challenges in visual impairment

Human social life begins with a radical shift in perspective, yet recent evidence suggests that human cognition is altercentric from the first years of life (Burkart & Southgate, 2025). This orientation has evolutionary roots. Unlike great ape infants, who cling to their mother's body and learn from her stance within an egocentric frame, human infants show a different developmental orientation. Due to an evolutionary adaptation toward face-to-face caregiving, they engage from a distance, adopting an other-centered perspective to perceive gestures and vocalizations (Braten, 2004). This shift in perspective enabled face-to-face interaction and mutual gaze. These behaviors lay the foundation for intersubjectivity, defined as the mutual attunement of actions and intentions that enables shared meaning and coordinated behavior between individuals (Bråten, 2007; Rizzolatti & Arbib, 1998; Trevarthen, 1998). Prenatal social interactions capture the congenital propensity for sociality in humans (Castiello et al., 2010). Newborns enter the world biologically prepared to engage socially, and they develop social understanding through reciprocal exchanges with caregivers (Striano & Reid, 2006; Trevarthen, 1979). Dyadic analysis techniques allow precise measurement of the timing of mother–infant contributions (Beebe & Stern, 1977; Stern, 1971). Such dynamics reveal the infant's active and immediate responsiveness to the adult's communicative intentions, defined as primary intersubjectivity (Trevarthen, 1979). A key transition in early development is from participating in face-to-face (dyadic) interactions to engaging in person–object–person (triadic) interactions. Triadic interactions involve two people in relation to some third external object, situation or event. When this shared orientation occurs, it constitutes an episode of joint attention. The conquest of this milestone is the birthplace of infants' sensitivity to others' mental lives s Additionally, it is linked to a series of developmental cascades consequences, such as language acquisition (Adamson et al., 2004), the conventional manipulation of objects (Bigelow et al., 2004), theory of mind development (Charman et al., 2000), and emotional regulation (Van Hecke et al., 2012).

Scientific results have shown that vision plays a crucial role in intersubjectivity processes. Eye contact helps infants identify when information is directed to them (Farroni et al., 2004), and typical caregiver–child interactions rely heavily on visual feedback, joint attention, and gaze following (Jongerius et al., 2020). When congenital or early-acquired visual impairment (VI) is present, these domains are profoundly affected (Grumi et al., 2021). Caregivers often struggle to interpret and respond to the child's communicative cues (Dale & Salt, 2007), while limited visual input and subtle alternative signals can hinder the emergence of intersubjectivity, conceived as the awareness of self and others (Damen et al., 2015). Since pioneering works (Fraiberg, 1968), research has shown that infants and children with congenital or early acquired VI may partially lack vision-dependent behaviors. The absence or disruption of visual signals such as eye contact, facial expression recognition, communicative gestures, and gaze monitoring (Lueck, 2008) can compromise emotional development and reduce opportunities for social engagement (Dale et al., 2014). These infants may also exhibit behaviors such as turning their head or body away from caregivers (Alfaro et al., 2021). A systematic review of the literature on parenting infants and children with VI revealed that discussions are largely dominated by the themes of intersubjectivity and joint attention (van den Broek et al., 2017b). Disruptions in early intersubjectivity may delay the development of joint attention, a key process for achieving shared focus with another person (Bigelow, 2003). Understanding the development of intersubjectivity in this context is essential to uncover how children with VI develop shared reference and the obstacles they face. Importantly, when provided with appropriate social experiences, children with VI can find alternative pathways to achieve co-reference. In sum, early intersubjectivity is not a single ability but a system of coupled perceptual, motor, affective, and contextual processes (Hobson, 2025). When vision is compromised, this system may reorganize rather than collapse, potentially recruiting alternative channels (e.g., touch, movement, posture, audition) to sustain shared engagement. This possibility motivates an ecological and dynamic framework aimed at empirically testing how intersubjectivity emerges and adapts over time.

1.2 Theoretical framework: ecological and dynamic system perspectives on "four-E" intersubjectivity

VI offers a privileged point of view for understanding intersubjectivity. When vision is reduced or absent, infants may need to reorganize how they engage with others, redistribute attentional resources, and rely on alternative sensory channels. This systemic reorganization reveals a broader developmental principle: intersubjectivity is not tied to a single modality but emerges from the continuous coupling of infants with their physical, social, and cultural environments (Ammaniti & Trentini, 2024). When one channel changes the entire system adapts to preserve engagement. This view naturally calls for theoretical models that conceptualize development as relational, adaptive, and multi-component.

Ecological and systems theories emphasize precisely this interdependence between individuals and their environments, showing that development does not unfold in isolation but emerges through continuous interaction (Bronfenbrenner, 1977). Dynamic Systems Theory complements this perspective by offering a flexible, time-dependent account of behavioral change, in which developmental phenomena arise from the interplay of multiple components within an open and adaptive system (Thelen, 2005; Thelen & Smith, 1994). Development unfolds across nested timescales, where changes at one level cascade into others: sensorimotor adjustments shape socio-relational dynamics, while relational experiences reorganize motor and emotional regulation (Masten & Cicchetti, 2010). These cascading effects are central to understanding intersubjectivity as a dynamic, embodied process. VI can therefore be seen not simply as a loss, but as a perturbation of the system that makes these adaptive couplings more evident.

From an ecological-functional standpoint, infants' behavior is adaptive and flexible, allowing them to select and modify actions in response to changes in body, environment, and task (Bernsteĭn, 1967; Lamb & Lerner, 2015). This adaptability is not only motoric but relational: actions are directed toward others, and perception–action loops sustain social connection. In the case of VI, relational dynamics should reorganize accordingly, relying on alternative strategies. Building on this foundation, the present thesis adopts the "four-E" perspective: Embodied, Embedded, Enculturated, and Enabling (Adolph et al., 2019) as a theoretical lens to examine the emergence of intersubjectivity depending on the way it emerges during development. Since intersubjectivity unfolds through development, the "four E" perspective provides a coherent framework for tracing how shared engagement becomes possible through bodily, environmental, cultural, and enabling conditions.

Development is Embodied. The morphology and functionality of the body shape the possibilities for interaction. Rapid physical growth during infancy prevents reliance on fixed associations, requiring constant recalibration (Adolph, 2020). Perception, action, and emotion form an integrated system (Thelen, 2005) that underpins both individual regulation and social engagement. In the perceptual domain, touch provides a sensory scaffold for constructing a sense of self (Bremner & Spence, 2017) and in VI this embodied grounding becomes even more evident: tactile, proprioceptive, and interoceptive cues take on a central role in differentiating self from other and in regulating social exchange. If development is embodied, then intersubjectivity, the capacity to share states and meanings with others, is likely to be grounded in the body, emerging through the dynamic coupling of perceptual and regulatory systems. When vision is compromised, the embodied nature of intersubjectivity becomes especially salient as interactions shift toward modalities that convey affective and regulatory information through direct bodily contact.

Development is Embedded. Perception and action are inseparable and situated. Information is actively generated through movement, rather than passively received (Gibson, 1979; Noë, 2004). From the very beginning, infants explore objects and people in their surroundings, searching for affordances for both physical and social action (Von Hofsten, 2009). Exploratory actions are movements intended to generate information (Gibson & Pick, 2000). Locomotion expands perceptual fields and opens new opportunities for social engagement; visual exploration and movement mutually reinforce one another, as seeing motivates action and action creates new conditions for seeing (Gibson & Schumuckler, 1989; Piaget & Cook, 1952). When visual input is reduced, exploration reorganizes: proximity, contact, and movement variability become functional strategies for managing uncertainty and maintaining connection. Evidence from clinical practice underscores that interventions should target mobility and social skills to improve participation and quality of life in VI children (Elsman et al., 2019). Because development is embedded, intersubjectivity is embedded as well: it unfolds through action in a situated environment, and when the availability of perceptual information changes, as in VI, patterns of exploration reorganize to sustain connection with others.

Development is Enculturated. Social information profoundly shapes action. Caregivers use facial expressions, vocalizations, and gestures to guide infants' and children's decisions, especially under uncertainty (Tamis-LeMonda et al., 2008). Cultural practices and designed environments scaffold these exchanges, embedding intersubjectivity in shared routines and norms. Congenital VI has been linked to significant delays in social interaction and referential communication (Tadić et al., 2009). Previous literature has focused on gaze as the primary attentional and referential cue (Abney et al., 2017; Land & Hayhoe, 2001), leaving body-based cues underexplored. Yet when vision is limited, alternative signals, such as head orientation, gain prominence as primary channels for initiating and maintaining interaction (Hammal & Cohn, 2014). Thus, if development is enculturated, intersubjectivity is likewise culturally mediated; when visual norms cannot be relied upon, caregivers recalibrate shared communicative practices, highlighting alternative bodily cues that support coordination.

Development is Enabling. Development does not simply add new abilities; it expands the space of perceptual and relational possibilities. Emerging motor skills extend infants' perceptual fields and create new opportunities for interpersonal engagement (Campos et al., 2000; Lerner et al., 2015). Beyond motor development, technologies such as sonification systems introduce technological affordances (Gibson, 1979) that reshape the environmental information. By translating movement into auditory feedback, these tools offer contingent information that enhances bodily awareness and spatial navigation. In absence of vision, this technology can serve a scaffolding function (Wood et al., 1976), operating within the zone of proximal development (Vygotsky, 1978), providing real-time cues that refine precision and synchrony, much like a caregiver's auditory guidance. If development expands the space of perceptual and relational possibilities, then intersubjectivity can also be enabled and extended through emerging skills and assistive technologies that open alternative channels for coordination and co-reference when visual guidance is limited.

1.3 Multisensory foundations of early intersubjectivity

The development of socio-emotional abilities depends on infants' capacity to integrate multisensory inputs, progressively acquire higher-order competences such as regulation and the attribution of meaning to others' actions (Rochat, 2014). These trajectories are rooted in early caregiver–infant interactions (Provenzi et al., 2022) and underpin mental health, well-being, and adaptive functioning across the lifespan (Sakurai, 2022). Within this context, caregivers' responsivity (accurate detection of the infant's communicative cues) and sensitivity (timely, attuned adjustment to those cues) shape infants' emerging ability to detect social signals and develop reliable expectations about others; importantly, optimal sensitivity reflects appropriately, not maximal, contingent responsiveness (Bornstein & Manian, 2013). Alongside these behavioral dimensions, parental embodied mentalizing adds a specifically multisensory layer to early intersubjectivity, as caregivers implicitly interpret intentions via movement, rhythm, posture, and touch and modulate their own bodily patterns accordingly, processes that scaffold infants' sense of agency (Shai & Belsky, 2011).

Critically, these early relational abilities rest on a multisensory foundation that both precedes and scaffolds later socio-cognitive development. From birth, functional tactile, auditory, interoceptive, and proprioceptive systems enable the detection of amodal properties, such as rhythm, temporal synchrony, and intensity; according to the principle of intersensory redundancy, information specified simultaneously in multiple sensory channels becomes particularly salient and learnable (Bahrick & Lickliter, 2000). At the same time, the relative weighting of simple temporal cues versus higher-level features changes with experience (Lewkowicz, 2014). Evidence from animal studies further illustrate this point: active engagement with sensorimotor contingencies is necessary for perceptual organization (Held & Hein, 1963), indicating that perceptual salience is shaped by experience, not only by stimulus redundancy.

Each sensory system contributes distinctively to early intersubjectivity. Touch regulates distress, establishes rhythmic co-regulation (Feldman, 2012) and support early sense of self (Bremner & Spence, 2017). Moreover, touch can actively scaffold learning in other modalities: when tactile sequences are temporally aligned with auditory patterns, infants use this intersensory redundancy to learn abstract tone structures, indicating that touch provides informative cues that enhance auditory statistical learning (Lew-Williams et al., 2019). Auditory input conveys prosodic cues essential for proto-conversational turn-taking and social orientation (Gratier et al., 2015). Proprioception continues to develop later in the development (Cowie et al., 2013) and contribute to bodily agency (Nardini et al., 2012).

In conditions of congenital or early-acquired VI, this multisensory architecture seems to reorganize. Rather than acting as simple compensatory channels, touch and audition often assume primary roles, leading to a re-calibration of sensory inputs (Gori, 2015). Because early socio-emotional exchanges rely heavily on amodal properties, rhythm, contingency, affective timing, many aspects of intersubjectivity remain accessible through alternative pathways. At the same time, the absence of visual signals reshapes how infants detect communicative bids and how caregivers interpret them (Grumi et al., 2021). As a result, early intersubjectivity emerges as an intersensory phenomenon, where coordination depends on the dynamic coupling of multiple channels; an organization made especially evident in VI, where the central role of bodily, auditory, and proprioceptive information in sustaining socio-relational development becomes particularly relevant.

1.4 Methodological consideration: from vision-centric models to multisensory approaches

The investigation of early social interactions during the developmental stages constitutes a critical research priority for the population of children with VI, as it enables the identification of strategies aiming to promote functional development and improving quality of life (Elsman et al., 2019). Among these interactions, the caregiver–child dyad has emerged as a domain of growing scientific interest (Provenzi et al., 2022; Urqueta Alfaro et al., 2021). Understanding how reduced visual input influences executive, motor, and relational functions is therefore essential for the development of theoretically grounded research models and evidence-based interventions (Gui et al., 2023). Given the high prevalence of social skill difficulties in this population and their cascading effects across developmental domains, clinical practice should integrate systematic assessment of social competencies and targeted instruction within early intervention programs (Caron et al., 2023). Family-centered approaches have further emphasized the importance of fostering interaction, intersubjectivity, and joint attention to strengthen caregiver–child relationships during infancy (van den Broek et al., 2017b).

Despite its theoretical and clinical relevance, research on intersubjectivity in children with VI remains markedly limited. This paucity of evidence is partly attributable to the low incidence of VI, particularly among children without additional disabilities. Within an alternative sensory scientific paradigm, recent studies have begun to investigate the role of non-visual cues in caregiver–infant exchanges. For instance, social touch by caregivers has been shown to regulate infants' emotional states (Cascio et al., 2019; Della Longa et al., 2021). A deeper understanding of alternative sensory modalities is critical for supporting parenting in VI contexts and informing clinical practices. However, well-established protocols for assessing attachment and relational security, such as the Strange Situation paradigm (Ainsworth et al., 1978), and traditional measures of joint attention (e.g., gaze alternation between partner and object) rely heavily on visual behaviors, making them poorly suited for this population (Alfaro et al., 2018). This reliance underscores a fundamental limitation: the assumption of vision as the primary modality for social engagement. Consequently, there is a pressing need not to merely replicate traditional paradigms but to reconceptualize assessment frameworks by integrating alternative sensory modalities, adopting metrics beyond gaze-based measures, and leveraging technological innovations. Such adaptations are essential to ensure ecological validity and inclusivity, while advancing both clinical practice and research in the domain of early social development in children with VI.

Nevertheless, social behaviors are inherently complex, dynamic, and context-dependent, which often leads researchers to impose experimental constraints and simplify behavioral phenomena. While such strategies may facilitate measurement, they risk neglecting essential aspects of the relational processes under investigation. Conversely, more ecological approaches, such as video-based coding, although valuable for capturing naturalistic interactions, are susceptible to increased observer bias and reduced reliability. To address these methodological challenges, technological innovation offers promising solutions. The growing emphasis on ecological validity in developmental research is likely to accelerate the design and implementation of advanced technological tools (Lee et al., 2018).

Moving beyond vision-centric assumptions requires the adoption of multisensory approaches and the development of technologies that leverage alternative modalities. Real-time sonification of movement emerges as a promising strategy to strengthen proprioception and, consequently, spatial cognition (Effenberg et al., 2016; Hermann et al., 2005; Nown et al., 2022). The implementation of automated sonification frameworks for navigation and interactive activities aligns with clinical evidence supporting early audio-motor training and enriched sensory environments as effective means to foster spatial orientation, body awareness, and motor development in children with VI (Cappagli & Finocchietti, 2017). Such technological solutions not only enhance ecological validity but also provide innovative tools for engaging and interactive rehabilitation programs.

1.5 Aim and objectives of the thesis

The aim of this thesis project has been to investigate how intersubjectivity develops when vision is impaired, exploring the adaptive role of alternative modalities, such as affective touch, proximity to the caregiver, head orientation, and movement sonification, in sustaining reciprocal engagement, shared attention, and co-reference. To achieve this, four objectives have been identified, which are described below.

The first objective focuses on *embodied intersubjectivity*, examining how interoceptive and tactile processes support relational dynamics through the physiological regulation of arousal in response to socially meaningful touch. This objective allows to explore whether and how visual deprivation modifies infants' sensitivity to the source of affective stimulation, highlighting touch as a foundational channel for early relational experience.

The second objective addresses *embedded intersubjectivity*, analyzing how exteroceptive input influences affective engagement and motor behavior. Adapting a protocol inspired by the Strange Situation paradigm to children with VI, this objective allows to investigate how unexpected sensory events modulate kinematic behavior, with particular attention to proximity to the caregiver, which regulates emotional activation, and to motor variability, reframed as a functional resource for maintaining interaction and environmental control.

The third objective concerns *enculturated intersubjectivity*, aiming to develop and validate a multimodal framework for capturing and quantifying joint attention during dyadic interactions. In children with VI, the development of joint attention becomes pivotal for rendering non-visual cues, such as head orientation, salient and interpretable. This objective allows to build a new method that uses head orientation as an alternative attentional index to gaze, combining clinical video coding, computer vision, and marker-based motion tracking techniques to detect episodes of joint attention in naturalistic caregiver–child exchanges.

Finally, the fourth objective explores *enabled intersubjectivity* through proprioception, introducing SONIA, an innovative sonification system that translates movement into sound. In the context of VI, enabling tools and protocols can augment access to contingencies that vision typically provides. By providing auditory feedback SONIA harnesses auditory and kinesthetic cues to strengthen bodily awareness, spatial navigation, and the sense of agency in the absence of vision. Rendering motor intent and timing audible can scaffold anticipation and alignment, offering a technologically mediated route to co-reference when visual guidance is limited.

Collectively, these objectives advance a framework that views intersubjectivity as a multidimensional process shaped by adaptive strategies in the absence of vision. Indeed, this thesis emphasizes the role of adaptation protocols and technological innovations in sustaining early social connection for guiding inclusive interventions.

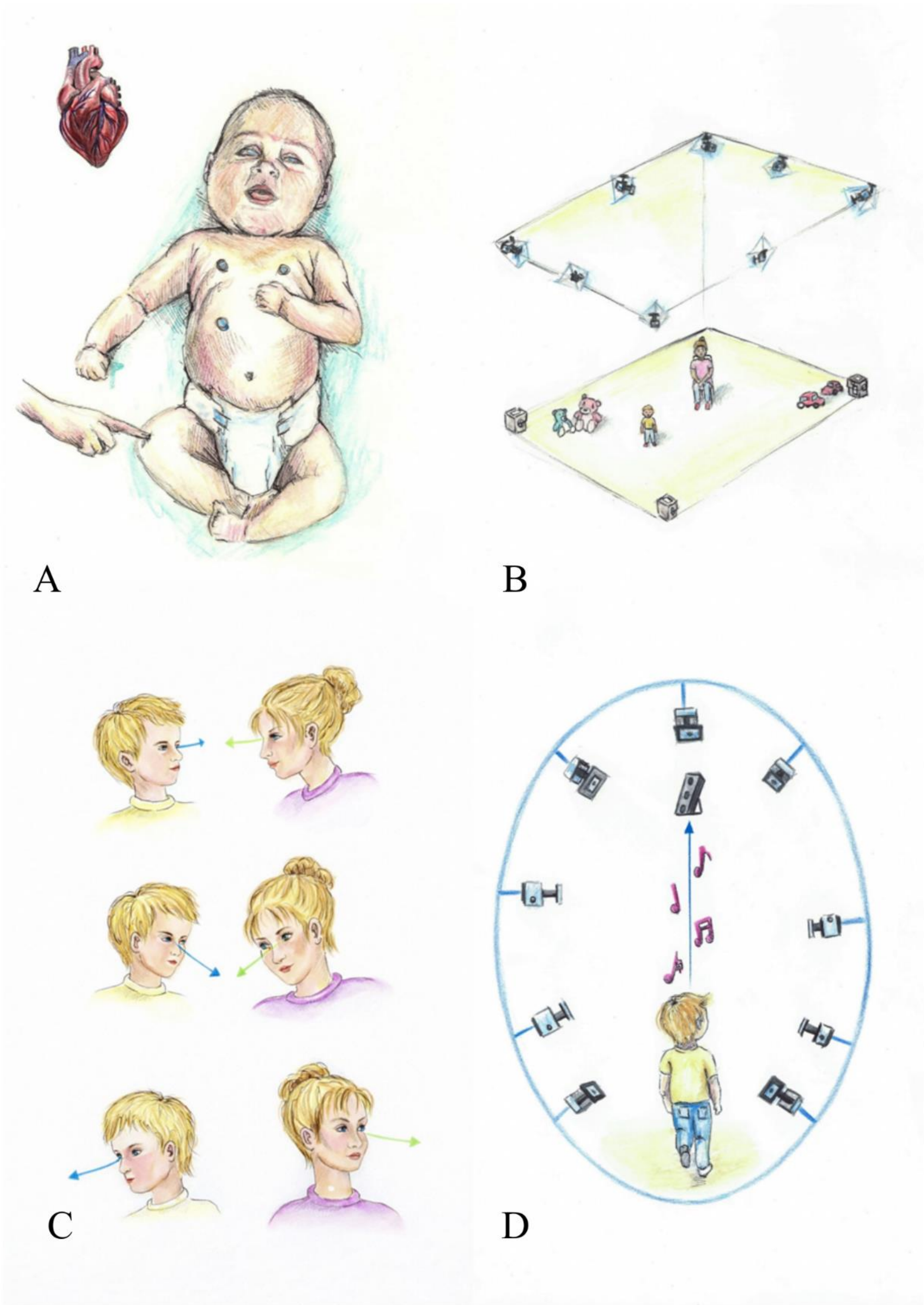

Figure 1.1 General Framework. (A) Embodied (B) Embedded (C) Enculturated (D) Enabling intersubjectivity.

Chapter 2

Embodied intersubjectivity as interoception.

The role of vision in caregiver-dependent physiological regulation through social touch

Summary

This chapter examines how affective touch operates as a channel for caregiver-dependent physiological regulation, exploring whether visual deprivation reorganizes the infant's sensitivity to the source of affective stimulation and reshapes the mechanisms of early relational attunement. A key challenge lies in understanding how the multisensory dynamics of caregiver–infant interaction is reorganized when one sensory channel is compromised. Building evidence that maternal and unfamiliar touch elicit distinct autonomic patterns, the proposed idea is to investigate whether visual deprivation alters this caregiver-specific modulation during infancy. Eighteen infants (9 VI, 9 typically developing (TD); 8–13 months) participated in a paradigm adapted from a previous work. Infants received affective stroking (0.3 cm/s) on the right shin while the person delivering the stimulation remained hidden. The parent or an unfamiliar experimenter seated once at a time in front of the infant. Heart rate (HR) index was computed as the relative change between stimulation and baseline. A mixed-model ANOVA revealed no significant main effects of Group or Caregiver Identity, but a significant interaction between the two. Post-hoc comparisons showed that TD infants exhibited reduced HR indices when interacting with the mother compared to a stranger, whereas VI infants did not show this differentiation. Moreover, VI infants displayed higher HR indices than TD infants in the mother condition. Findings suggest that visual deprivation might reorganizes caregiver-specific physiological tuning, indicating that the visual channel plays a fundamental role in shaping affective development.

2.1 Introduction

2.1.1 The physiological roots of affection

Describing a baby inevitably means describing a baby and someone else (Winnicott, 1948). This quote from Winnicott illustrates the crucial importance of relations for our perceptual and cognitive development. Humans appear to have an inherent sensitivity to engage with their social environment (Adolphs, 1999; Meltzoff & Brooks, 2001). Indeed, some authors emphasized broader innate needs related to contact, movement, and temperature (Winnicott, 1964). Nonetheless, the prevailing psychological theoretical view has long considered affection a secondary drive, formed through repeated associations between the mother and the reduction of primary biological needs, particularly hunger and thirst (Klein, 1957). In contrast, others emphasized broader innate needs related to contact, movement, and temperature. Landmark experiment (Harlow & Zimmermann, 1958) explored human–animal parallels through the dual mother surrogate condition: a cloth mother and a wire mother that provide food. Infant monkeys preferred the cloth mother, demonstrating that infant-mother relation in the form of contact comfort is a dominant factor in the development of affectional responses, whereas infant-mother relation in the form of nutritional needs plays a negligible role. During development, the spontaneous engagement in social relations functions as a survival strategy, ensuring access to resources essential for growth, protection, and reproduction (Dunbar & Shultz, 2007). In social species, individuals regulate one another's physiological processes, making survival dependent on social bonds. Indeed, caregivers often act as "hidden regulators" of infant physiology (Hofer, 1994). As human infants, we are immersed in continuous streams of interconnected signals from caregivers, including speech, gestures, eye gaze, and touch (Abu-Zhaya et al., 2017; Frith & Frith, 2007). In particular, touch represents a crucial psychophysiological component in the ontogeny of intersubjectivity: early bodily experiences of touch, beginning in gestation, contribute to the emergence of a social self from a psychophysiological perspective (Di Plinio et al., 2022). The sense of touch is the first sensory modality to develop during gestation and is distributed across the entire body (Jonas, 1973; Ratcliffe, 2013). As the earliest developing sense (Maurer & Maurer, 1988; Montagu, 1971), touch provides a sensory scaffold for perceiving one's own body and constructing a sense of self (Bremner & Spence, 2017). Identity and alterity, jointly with reciprocity, start to coexist in intersubjectivity during the earliest tactile interactions with infants (Gallese & Ebisch, 2013). Regarding this, touch is prominent in infant-caregiver dyadic interactions (Feldman et al., 2010; Ferber et al., 2008; Herrera et al., 2004) and has been shown to play a role in directing infants' attention, regulating arousal levels, and reducing distress (Hertenstein, 2002; Jean & Stack, 2012).

2.1.2 The role of affective touch in neuro-cognitive development

Experiencing touch, both in giving and receiving, plays a vital role in shaping neuro-cognitive growth (Brauer et al., 2016; Simpson et al., 2019). Research has revealed that the hairy skin of mammals contains C-tactile (CT) afferent fibers, specialized receptors that respond optimally to gentle, slow stroking at a force, velocity, and temperature typical of social grooming among primates (Morrison et al., 2010). Activation of these fibers is consistently perceived as affectively pleasant. These hedonic properties of touch are mediated by unmyelinated, low-threshold mechanoreceptors, CT fibers (Löken et al., 2009), which project to limbic regions such as the insular cortex, a hub for emotional and interoceptive processing (McGlone et al., 2014; Vallbo et al., 1999) . The parallels between CT-mediated touch and oxytocin release, both associated with reduced physiological arousal, pleasant feelings, and prosocial interaction, suggest that CT fibers may play a key role in facilitating endogenous oxytocin release during affiliative and nurturing touch (Walker et al., 2017) . This discovery has highlighted human skin as a social organ (Morrison et al., 2010), making the concept of affective touch particularly important for early development. From birth, affective touch evolves from a passive sensory experience into a reciprocal interaction between caregiver and infant. Affective touch can also arise from actively delivering CT-optimal stimulation, even through glabrous skin, producing pleasant sensations for both giver and receiver (Gentsch et al., 2015). Social touch is essential for healthy physical and psychological development from birth (Carlson & Earls, 1997; Field, 2010). The ability to perceive and respond to pleasant touch is vital early in life because of its role in affiliation, bonding, and biobehavioral synchrony, particularly between caregivers and infants (Feldman, 2012; Ferber et al., 2008). At 9 months, CT-targeted touch reduces heart rate and increases engagement (Fairhurst et al., 2014), with responses modulated by the source of touch (Aguirre et al., 2019).

Evidence for a functioning pleasant-touch system in infants provides a foundation for future research aimed at systematically examining its physiological mechanisms and its influence on social and cognitive development, both typical and atypical. Indeed, manipulating affective touch offers a promising approach to studying interoceptive signaling in humans (Cascio et al., 2019). Within this framework, an important challenge is understanding how the multisensory nature of caregiver–infant interaction is reorganized when one sensory channel is compromised. Visual deprivation might hamper social interactions because it deprives individuals of a crucial source of interaction, namely visual feedback, potentially impacting bonding and biobehavioral infant-mother synchrony.

The present work builds on previous paradigms (Aguirre et al., 2019) to explore whether the absence of vision impacts the discrimination and integration of affective touch, with potential consequences for bonding and biobehavioral synchrony.

2.2 Method

2.2.1 Participants

This study has been carried out as part of a collaboration between the Center of Child Neuro-Ophthalmology at IRCCS Mondino Foundation (Pavia, Italy), and Unit for Visually Impaired People at Istituto Italiano di Tecnologia, (Genoa, Italy). The experimental protocol was approved by the local Ethics Committee (Fondazione IRCCS Policlinico San Matteo, Pavia, Italy) and all parents participants gave their written informed consent under the Declaration of Helsinki.

The inclusion criteria was a) VI of peripheral origin (with an involvement of the pregeniculate visual pathway). Exclusion criteria were a) presence of neuromotor disorders, central nervous system involvement, and/or or chronic comorbidities; c) presence of additional sensory deficits.

Participants were recruited from the clinical population in care at the Centre of Child Neurophthalmology, IRCCS Mondino Foundation, Pavia (Italy). According to the criteria described above, 9 Typically Developed (TD) and 9 Visually Impaired (VI) infants from 8- to 13-month-old were tested. By 8–9 months of age, infants typically display a well-established aversive reaction to strangers (Sroufe, 1977). In addition, evidence suggests that before 8 months old, the cerebral processing of affective touch might still be immature (Kida & Shinohara, 2013; Pirazzoli et al., 2019).

2.2.2 Materials

Two identical synthetic fiber brushes were employed in the procedure. One brush was positioned behind a curtain and used to deliver tactile stimulation, while the other was held by an adult (either the participant's parent or an unfamiliar experimenter) acting as the potential visible source of touch. Heart rate was monitored using a three-lead electrocardiogram system (Biopac MP36) with Biopac EL104 electrodes. Two electrodes were placed below the clavicles, and a third was positioned on the left floating rib. The session was recorded with a single camera operating at 30 frames per second. Participants were seated in an infant chair. A large plastic tray combined with opaque fabric panels attached to the chair prevented visual access to the person delivering the tactile stimulation. Throughout the procedure, the individual applying the strokes (Experimenter 1) remained behind an opaque curtain positioned to the participant's right (orientation based on the participant's viewpoint). A small opening in the curtain allowed Experimenter 1 to brush the participant's right shin. In addition to Experimenter 1, two other adults were present: the participant's mother and a second experimenter (Experimenter 2). A pre-recorded audio file was used to mark the phases within each experimental block and to cue position changes between the mother and Experimenter 2.

Figure 2.1: Left: Setup. Right: electrode placement. Left: Schematic representation of the participant's setup, including the infant, the unfamiliar experimenter (Experimenter 2), the mother and the hidden experimenter (Experimenter 1). A plastic tray prevented the infant from seeing who applied the tactile stimulation, while a curtain divided the space and concealed Experimenter 1 from the infant's view. **Right:** Illustration of electrode placement and depiction of the stimulated area on the infant's leg.

2.2.3 Procedure

The procedure comprises four testing blocks, each composed by four phases (resting, brush, stimulating, resting) (Figure 2.2). Each testing block began with a 10-second positioning phase, during which the two adults (parent or Experimenter 2) moved to their assigned seats: one positioned in front of the participant and the other to the left, both approximately 60 cm from the infant chair. The adult seated in front placed the paintbrush next to the participant's right leg without making contact (brush-down position).

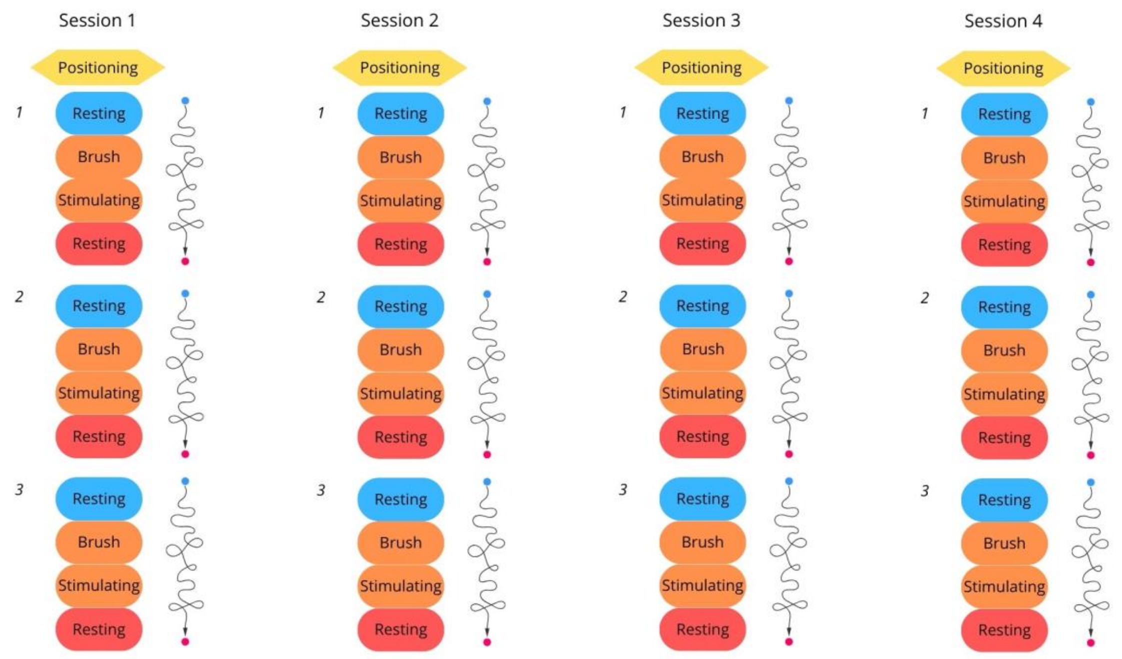

Figure 2.2: Overview of the experimental flow. The experiment consists of four sessions. Each session begins with the positioning of the parent and the stranger, who switch places across sessions. Within each session, there are three blocks, and each block includes four phases.

The resting phase requires the adult in front of the infant to maintain the brush in the brush-down position without delivering tactile stimulation. During the brush phase (5 s), the adult seated in front of the infant held the brush and guided the infant's hand to make contact with it, allowing tactile exploration of the object. The adult also spoke to the infant to signal their presence. This is the phase for which a first adaptation was needed with respect to the original paradigm (Aguirre et al., 2019) where the presence of the brush and recognition of the adult (parent vs stranger) were made evident just through visual feedback, namely letting the infant seeing the brush in the parent's hand and the adult in front of him.

After approximately 2 s of contact, the brush was returned to the brush-down position next to the participant's leg. This presentation ensured that the infant could perceive the brush in the adult's hand. The stimulating phase (10 s) replicated the resting setup except that Experimenter 1, concealed behind the curtain, stroked the participant's right shin at a velocity of 3 cm/s. Strokes were applied along an axis parallel to the tibia over a 3 cm area, following a previous protocol (Tuulari et al., 2019).

Heart rate was recorded during each stimulation phase. To estimate baseline cardiac activity, heart rate was also measured during the 10-second resting period preceding the brush presentation and the 10-second resting period following the stimulation offset. For each trial, baseline heart rate was calculated as the average of these two resting intervals. At the start of the second, third, and fourth blocks, the parent and the unfamiliar adult exchanged positions, ensuring that each adult sat in front of the participant for two blocks. The identity of the adult seated in front during the first block was counterbalanced across participants. A prerecorded audio track provided cues for brush presentation and for position changes during the positioning phases.

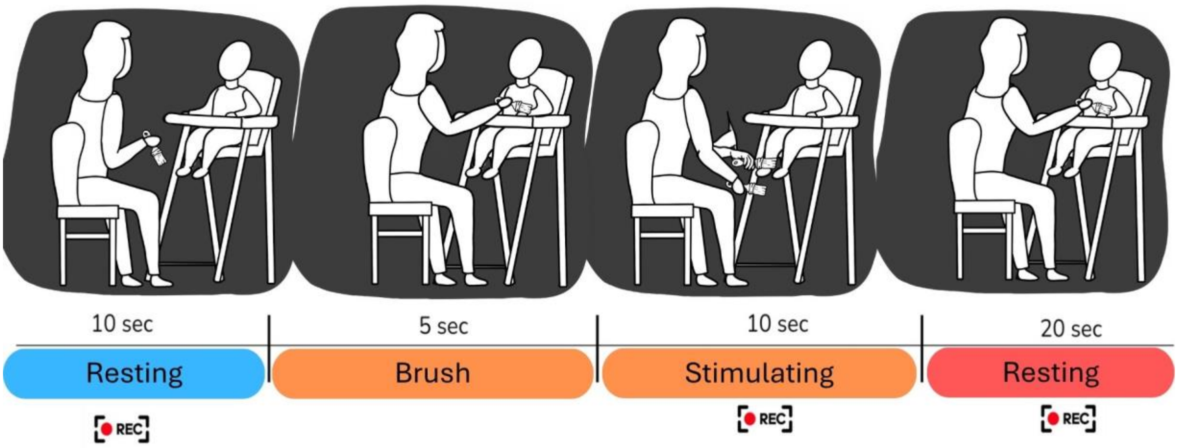

Figure 2.3: Experimental procedure. Each block included four phases: Resting (pre-stimulation), Brush, Stimulating, and Resting (post-stimulation). The figure illustrates the seating arrangement of the infant, the parent or experimenter 2, and the experimenter 1 behind the curtain. The adult seated in front of the infant acted as the potential visible source of touch, while experimenter 1 delivered tactile stimulation to the infant's right leg through an opening in the curtain.

2.2.4 Data analysis

ECG signals were acquired using a Biopac system equipped with a hardware band-pass filter (0.05–150 Hz) and a sampling rate of 2000 Hz. Heart rate was then calculated in beats per minute (BPM) from the raw ECG signal. For each condition (parent vs. stranger), mean heart rate was computed separately for the stimulation phase (0–10 s from stroking onset) and for baseline periods (10 s preceding brush presentation and 10 s following stimulation offset). These values were averaged across trials within each identity condition. Finally, the heart rate change ratio (Hr) was derived for each participant as the difference between stimulation and baseline mean heart rates divided by the baseline mean heart rate, using a formula of a previous study (Loggia et al., 2011).

2.3 Results

2.2.5 Findings

Analyses confirmed that the data did not violate assumptions of normality (Shapiro–Wilk tests, $p > .05$) or homoscedasticity (Levene's test, $p > .05$). A mixed-model ANOVA with Condition (Stranger vs. Mother) as a within-subject factor and Group (VI vs. TD) as a between-subject factor revealed no main effect of Group, $F(1, 16) = 0.05$, $p = .827$, $\eta^2 = .002$, and no main effect of Condition, $F(1, 16) = 0.20$, $p = .660$, $\eta^2 = .005$. There was a significant Group × Condition interaction, $F(1, 16) = 8.05$, $p = .012$, $\eta^2 = .172$.

Post-hoc comparisons indicated that within the VI group, HR index did not differ between Stranger and Mother ($p = .111$), whereas within the TD group, HR index was significantly lower for Mother compared to Stranger ($p = .034$). Between groups, a marginally non-significant difference emerged in the Stranger condition, with VI children showing a slightly lower HR index than TD children (Mean difference = –0.0153, SE = 0.0106, $p = .168$), while in the Mother condition, VI children exhibited a significantly higher HR index than TD children (Mean difference = 0.0184, SE = 0.0076, $p = .028$).

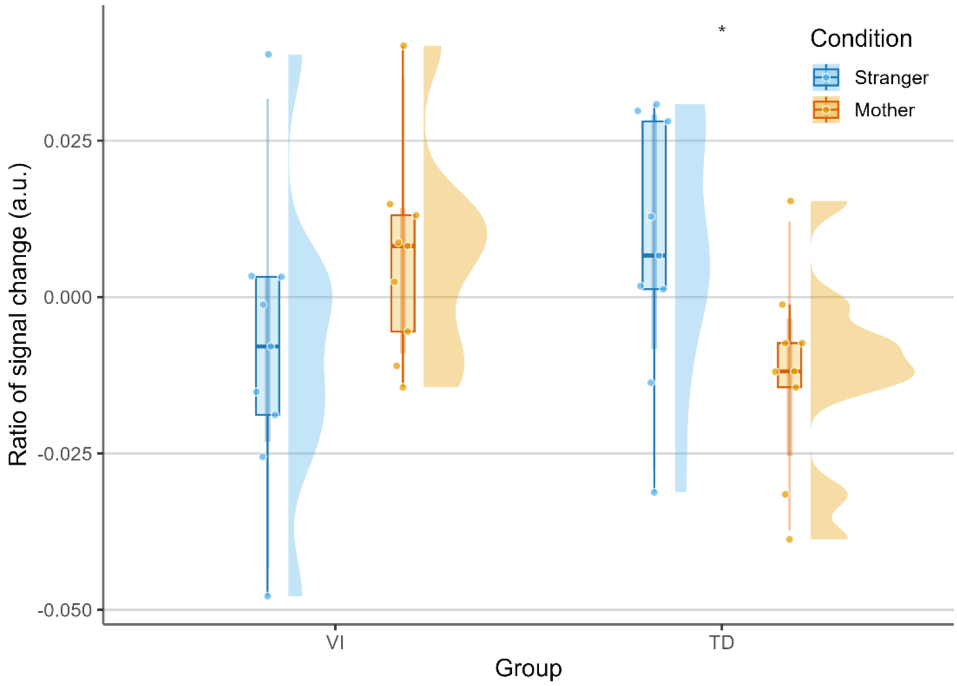

Figure 2.4: Raincloud plot showing the distribution of individual HR indices across conditions (Stranger vs. Mother) for infants with VI and TD peers. Half-violins depict density estimates, boxplots summarize central tendency and variability, horizontal bars indicate group means, and jittered points represent individual participants.

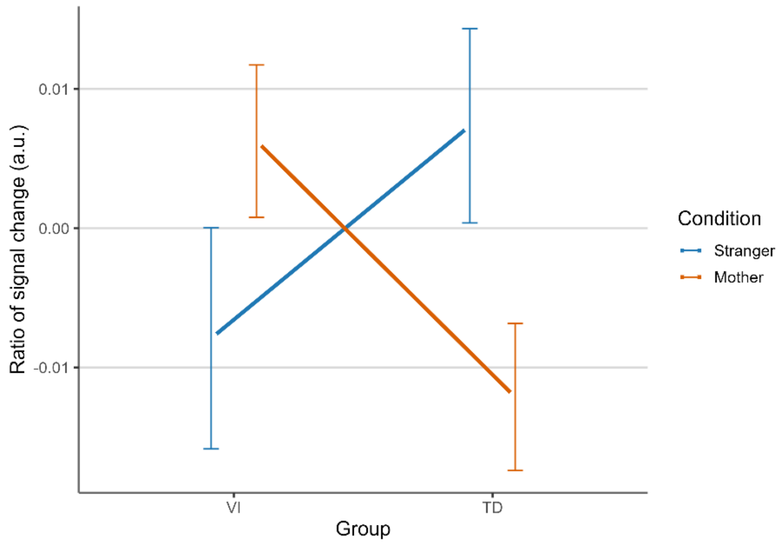

Figure 2.5: Interaction plot of mean heart rate (HR) indices across conditions (Stranger vs. Mother) for infants with VI and TD peers. Lines represent group means, with error bars indicating ±1 SE.

2.4 Discussion

The present study investigated whether caregiver identity influences affective touch processing in infants with and without visual disability. The rationale was to determine the extent to which visual feedback guides the development of affective touch. During the procedure, tactile stimulation was applied to the infant's shin by a hidden experimenter while heart rate was continuously monitored to assess autonomic regulation. The 'social' component of the touch was manipulated by positioning either a parent (socially relevant) or a stranger (socially non-relevant) in front of the infant. This setup ensured that the visible and audible adult could plausibly be perceived as the source of stimulation, making it possible to test whether caregiver identity modulated infants' physiological responses under conditions of visual deprivation. The main hypothesis was that vision plays a crucial role in early social interactions. Accordingly, VI infants were expected to show altered affective touch processing, specifically a lack of differential physiological response when tactile stimulation occurred in the presence of the mother versus a stranger.

Results did not show significant main effects of Group or Caregiver Identity. However, the significant Group and Caregiver Identity interaction suggests a divergence in how TD and VI infants modulate their autonomic responses to social touch. Post-hoc comparisons indicated that TD infants showed a significantly lower heart rate index when touch was delivered by the mother compared to the stranger, whereas this difference was not observed in VI infants. Additionally, VI infants exhibited a higher HR index than TD infants in the mother condition. These patterns must be interpreted with caution: the study was likely to increase the risk of Type II error due to the small sample size. In addition, heart rate measures in infancy are characterized by substantial interindividual variability, which may have further limited sensitivity to condition-specific effects. Rather than providing confirmatory evidence, the observed trends might help generate hypotheses about caregiver-specific autonomic tuning and its possible alteration in the context of visual deprivation. Future studies with larger samples and improved power will be necessary to disentangle group differences from variability-related noise and to clarify the robustness of caregiver-related autonomic effects.

Consistent with prior literature, tactile communication is pervasive in early caregiver–infant interactions: approximately 65% of face-to-face exchanges involve touch, which is consistently associated with reductions in behavioral and physiological stress (Feldman et al., 2010; Stack & Muir, 1990). In this context, maternal touch functions as a powerful regulator of infant physiology, modulating temperature, heart rate, sleep, arousal, and cortisol levels through skin-to-skin contact (Feldman et al., 2002; Feldman et al., 2011). The heightened cardiac response to maternal presence observed in VI infants may therefore reflect an adaptive reliance on tactile cues to compensate for the absence of visual information.

Moreover, parallels can be drawn with research on preterm infants, where tactile and kinesthetic stimulation promotes growth and physiological regulation, improving weight gain, feeding (White & Labarba, 1976) and responses to painful stimuli (Maitre et al., 2017). CT-optimal stimulation in preterm infants has been shown to positively affect heart rate and oxygen saturation compared to static touch (Manzotti et al., 2019). These findings have informed clinical practices such as massage therapy, kangaroo care, and osteopathic manipulative treatment, all of which leverage the regulatory power of social touch (Als & B. McAnulty, 2011; Boundy et al., 2016; Cerritelli et al., 2013; Gallace & Spence, 2016).

2.5 Conclusion

Embodied intersubjectivity: disruption of caregiver-specific autonomic tuning under visual deprivation

Taken together, the present results contribute to this body of evidence by showing that, in the absence of visual cues, infants with VI do not exhibit the expected cardiac deceleration when the apparent source of social touch is the mother. The contrast between VI and TD infants highlights the importance of vision in shaping early affective responses, while also underscoring the compensatory potential of tactile communication in supporting regulation. Importantly, these findings align with the embodied framework: if early intersubjectivity emerges through the dynamic coupling of perceptual and regulatory systems, then visual deprivation is expected to reorganize how infants rely on bodily cues to interpret socially meaningful contact. The divergence in caregiver-specific autonomic modulation observed here is consistent with this view, suggesting that visual constraints may subtly alter caregiver-specific physiological tuning. These insights not only advance understanding of multisensory reorganization in infancy but also carry implications for interventions aimed at fostering bonding and physiological stability in populations with sensory impairments, where touch may function as a foundational channel for early relational experience.

Chapter 3

Embedded intersubjectivity as exteroception.

Rethinking the strange situation protocol for children with visual impairment

This chapter has been published in *Research in Developmental Disabilities* (Guarischi et al., 2025).

Guarischi, M., Montagnani, E., Catalano, G., Saligari, E., Signorini, S., & Gori, M. (2025). From motion to interaction: How multisensory information shapes motor behaviors in children with visual impairment. Research in Developmental Disabilities, 159, 104956. doi: 10.1016/j.ridd.2025.104956

Summary

This chapter investigates how multisensory events modulate affective engagement and motor variability in children with VI by adapting the Strange Situation paradigm to examine how environmental cues reorganize caregiver–child interaction dynamics. To address this question, the study analyzed children's motor trajectory variability and their proximity to caregivers in playroom conditions with and without multisensory stimulation. Data were collected using a Motion Capture System within an ecological playroom environment.

The Strange Situation procedure was reconceptualized for this population by placing three TechARMs in the corners of the playroom; these devices delivered audiovisual cues and acted as a substitute for the stranger. When multisensory cues were introduced, children with VI displayed greater variability in their motor trajectories and maintained a shorter distance from caregivers. These signals appeared unusual to the children and influenced kinematic patterns, suggesting that environmental unpredictability shaped how attention and motor control were allocated.

Overall, the findings contribute to current evidence on motor and social interaction in children with VI, showing how multisensory inputs can either scaffold or disrupt social engagement and exploratory behavior. Enhancing ecological validity in such paradigms may support further advances in the design and application of technological tools in both research and clinical settings.

3.1 Introduction

3.1.1 Independent locomotion, visual input and the foundations of early social interaction

Independent locomotion, whether by crawling, shuffling, or walking begins an extraordinary psychological reorganization that enables infants and children to investigate their social environment (Adolph, 2008). Indeed, a stable postural base creates new possibilities for acquiring knowledge and acting on the world (Adolph & Franchak, 2017). Consequently, the capability to plan and guide motor actions needs a person to perceive and understand the neighboring space, whilst social factors have the potential to either peak or diminish motor actions. In the enterprise of independent locomotion vision serves multiple roles, including maintaining balance, avoiding collisions, steering, navigating and adapting to changes in the surface of the ground (Adolph, 2020). Indeed, visual perception shapes motor behavior by giving feedback on the effects of previous movements, offering clues about the next steps to take (Adolph & Franchak, 2017). This way, vision connects actions to recent as well as past experiences, offering a guide for future activities (Adolph, 2020). Stimuli inside the visual field, especially those with a social value, provide an incentive to move. In fact, vision represents a key component in the development and promotion of early parent-child interaction. Consequently, typically early human interactions mostly rely on visual feedback, joint attention and gaze following (Jongerius et al., 2020), domains impaired when an early acquired or congenital VI is present (Grumi et al., 2021). Thus, the limited visual input from caregivers, combined with unique signals that might go unnoticed by them, could contribute to delays in the development of intersubjectivity. Specifically, missing or disrupted visual signals like eye contact, recognition of facial expression, communicative gestures, gaze monitoring, and visual imitation could compromise the emotional development of children with VI, limiting opportunities to engage in social relationships (Dale et al., 2014). Therefore, congenital VI is frequently associated with atypical developmental trajectories, particularly in social interaction skills, which may experience setbacks during the second and third years of life (Vervloed et al., 2020). Consequently, children with VI are at risk of negative social experiences, such as peer rejection and reduced accessibility to social activities (Chaumet-Riffaud et al., 2017; DeCarlo et al., 2012; Rainey et al., 2016). Indeed, preschoolers with VI often engage more in solitary activities and have fewer interactions with peers compared to children with typical vision (Cappagli et al., 2018; Caron et al., 2021). Moreover, this effect persists across the lifespan, as adults with VI have been found to exhibit difficulties in social evaluations (Ferrari et al., 2017), reduced participation in leisure activities (Salminen & Karhula, 2014), and higher levels of loneliness (Brunes et al., 2019).

Holding these premises, in the presence of developmental VI, research on early social interactions is mandatory, in order to detect possible strategies to promote patients' overall functioning and quality of life (Elsman et al., 2019). Particularly, the interaction between children with VI and parents represents a topic of growing interest (Gui et al., 2023; Provenzi et al., 2022; Urqueta Alfaro et al., 2021). As Lee (2018) pointed out, the growing emphasis by researchers on enhancing ecological validity could drive progress in the development and application of technological devices. Indeed, the use of new technologies in VI children rehabilitation represents an upcoming challenge (Gori et al., 2016). Recently, multisensory devices resulted efficacious in improving mobility and spatial cognition in such patients (Cappagli et al., 2019). Moreover, technologically augmented toys with auditory cues have been shown to facilitate social play by increasing shared attention between children with VI and their sighted peers (Verver et al., 2019, 2020). Accordingly, a recent review (Elsman et al., 2019) on the effectiveness of rehabilitation interventions for individuals with VI underscored the need for future research aimed at enhancing participation and quality of life, with a particular focus on improving mobility and social skills.

Building on these findings, this study aimed to investigate whether the use of multisensory stimuli could impact motor exploration and the relational dynamics between children with VI and their parents. For this, a Vicon motion Capture System and the TechARMs, technological devices that could deliver audio-visual stimuli (Schiatti et al., 2020), were used. With this setting, two motor parameters were investigated under ecological play: 1) the variability of children's motor trajectory with respect to a common point (also referred to as space origin), and 2) the distance between children and caregivers. Starting from a previous work (Budman et al., 2019), the setup was adapted to the ecological interaction with the attachment figure. This appears to be the first study that investigates spontaneous motor behaviors of children with VI from kinematic analysis and during playroom activities.

3.2 Method

3.2.1 Participants

This study has been carried out as part of a collaboration between the Center of Child Neuro-Ophthalmology at IRCCS Mondino Foundation (Pavia, Italy), and the Unit for Visually Impaired People at Istituto Italiano di Tecnologia, (Genoa, Italy). The experimental protocol was approved by the local Ethics Committee (Fondazione IRCCS Policlinico San Matteo, Pavia, Italy) and all parents participants gave their written informed consent under the Declaration of Helsinki. Children were recruited from the clinical population in care at the Centre of Child Neurophthalmology, IRCCS Mondino Foundation, Pavia (Italy). The inclusion criteria were: a) VI of peripheral origin (with an involvement of the pregeniculate visual pathway); b) visual acuity <0.5 logmar scale, and c) the ability of walking confidently for at least one year (i.e., being able to take more than 20 steps without falling or losing balance, pick up toys while walking, navigating around objects and interact with caregivers while walking) (Price et al., 2022). Visual acuity was evaluated under maximum refractive correction with test suitable for patient's age and cooperation using Teller Acuity Cards (Teller et al., 1986), Lea symbols or letter optotypes (Hyvärinen et al., 1980). The visual acuity was defined as "reduced" according to normative data (Donahue et al., 2016). Exclusion criteria were: a) presence of developmental delay, based on standardized evaluations chosen according to the age and the degree of VI detected with Griffiths Scales of Child development (Green et al., 2016) and the Reynell-Zinkin scale (Reynell & Zinkin, 1975); b) presence of neuromotor disorders, central nervous system involvement, and/or or chronic comorbidities; c) presence of sensory deficits and psychiatric disorders. According to the criteria described above, 11 children were selected to take part in the present study. In Supplementary Table 3.1, demographics information related to participants was reported. In Supplementary Table 3.2, additional clinical data was reported for each participant.

3.2.2 Materials

The experiment was conducted within the Joint Lab of the IRCCS Mondino Foundation, Pavia. The experimental procedure was explained to the parents of the participants prior to testing; a playground set up has been created where children could play freely as they normally do in their home/childcare environment. The set up consisted of a modular square carpet (approximately 2.8x2.8m) that was surrounded by a safety gate to ensure that children could not exit the motion capture cameras' view. The space was also provided with toys and soft cushions so that children could play comfortably (Figure 3.1).

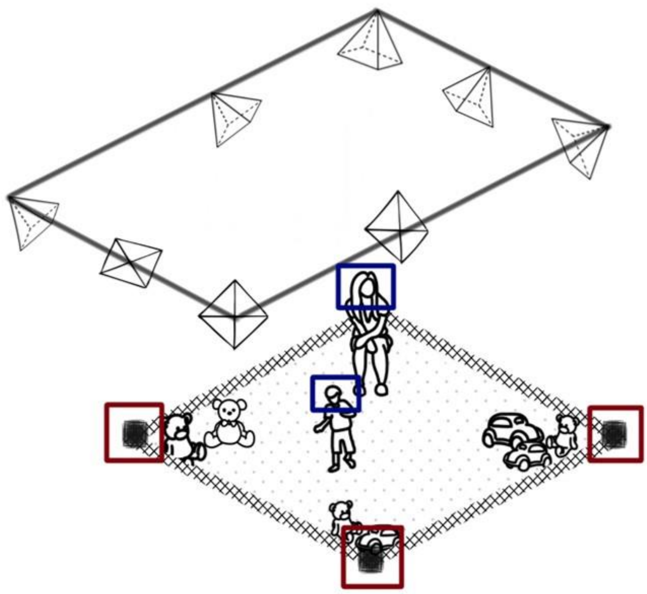

Figure 3.1: Experimental setup. The Vicon System consisting of eight near-infrared cameras and the TechARMs at the corners of the carpet, near three areas with various toys. The attachment figure is sitting in a corner, and the child moves inside the play area. For an enlargement of the red and blue boxes, refer to Figures 2 and 3.

The Vicon System (Vicon Motion Systems Ltd. UK) with eight near-infrared cameras (Vero cameras, 100 Hz) was used to reconstruct the motor trajectory of the reflective markers that were placed at specific anatomical landmarks. In particular, both children and caregivers wore head bands with three markers that were placed on the right and left temples, as well as on the middle forehead (Figure 3.2). Multisensory stimuli during playground activities were given with the TechARMS, which were placed at three corners of the square space. The TechARM (Figure 3.3) is a portable device comprising sensors and actuators to enable visual, haptic and auditory interaction. In this specific procedure, this wireless device was used for visuo-audio stimulation.

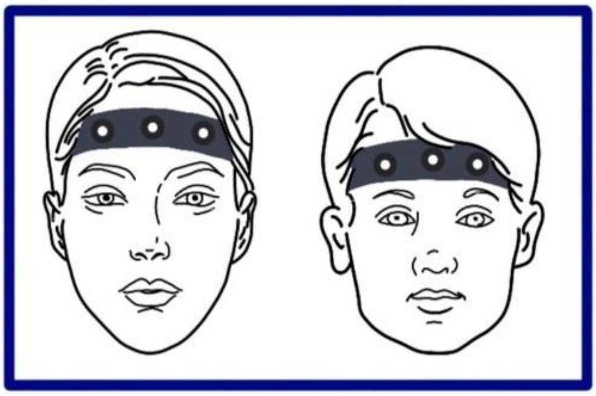

Figure 3.2: Configuration of markers, for the attachment figure and the child.

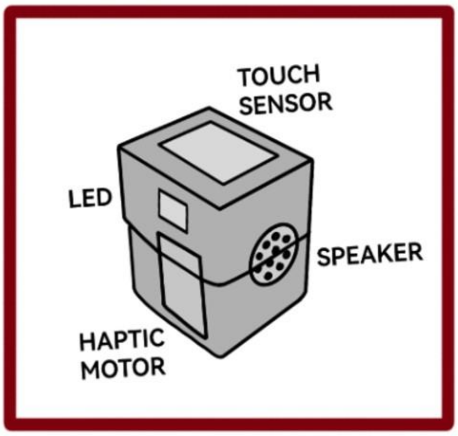

Figure 3.3: Graphic representation of the TechARM, composed of LEDs, touch sensors and speakers.

3.2.3 Procedures

Participants and their caregivers were welcomed into the testing room and introduced to the playground environment. At this stage, informed consent was obtained from caregivers and demographic information was collected. Caregivers and children were then familiarized with the setting and informed that, during the session, visual and auditory stimuli would be presented through small cubes (techARMs) placed within the play area.

Caregivers were instructed to interact naturally with their child, while avoiding actively encouraging locomotion or directing the child's behavior. Children were free to explore the environment spontaneously.

The experiment consisted of 13 trials, each of one minute, with three types of conditions:

- Baseline condition (three trials): to promote familiarization with playground space and detect spontaneous motor behavior.
- Multisensory condition (five trials): the techARMs were switched "ON" remotely from a custom written Python script, providing visual (green light) and audio stimuli (frog sound) for three seconds every 1.5 seconds.
- Silent condition (five trials): no sensory stimuli were provided, as the techARMs were switched "OFF".

The multisensory stimulation included both auditory and visual components to reflect the sensory experiences of the target population. Although participants had VI, all enrolled children possessed some degree of functional residual vision, which is actively exploited in daily activities and rehabilitative contexts. Accordingly, the inclusion of visual stimulation was intended to maximize ecological validity and to engage the participants' available sensory resources.

The presentation of multisensory and silent conditions was alternated in order to minimize potential effects of progressive familiarization with the experimental environment. This procedure was designed to introduce controlled variations in sensory input while keeping the caregiver continuously present, in a static and fixed position. The experimental protocol was conceptually inspired by principles underlying the Strange Situation paradigm (Ainsworth et al., 1978), particularly the use of structured environmental changes to elicit regulatory and proximity-related behaviors. Importantly, the present procedure does not constitute a direct adaptation of the classical Strange Situation, as it did not include caregiver separation phases nor the physical presence of a stranger. Within this framework, the baseline condition was intended to capture spontaneous exploratory behavior in a familiar context. The multisensory condition introduced an unusual combination of sensory cues, operationalized as an environmental perturbation rather than as a substitute for a social stranger, aimed at eliciting regulatory responses to novelty. The subsequent silent condition followed the perturbation and was designed to observe changes in caregiver-directed proximity behavior. Given the lack of explicit separation–reunion episodes, the present protocol cannot be interpreted in terms of classical attachment classifications. Instead, it should be considered an exploratory, attachment-informed procedure aimed at investigating motor and proximity-related strategies under varying sensory contexts.

3.2.4 Data processing

Space calibration was performed in the Vicon Nexus Software before data capture started, according to the system specifications. Then, a 3D model was created using the head as a single kinematic segment where three markers were identified (left, central and right foreheads), for both parents and children. Following data pre-processing in Nexus Software, the x/y coordinates of the three forehead markers were extracted from the .c3d files from a custom written Matlab script (v R2023b, The MathWorks, Natick, USA). These coordinates were then filtered using a low-pass Butterworth filter with a 6 Hz cutoff to remove any residual noise in the data (Winter et al., 1974). For the present study, only the motion trajectories from the central forehead marker of both children and parents were utilized. From the filtered forehead trajectories, two kinematic measures were computed. First, the variability of the children's motion trajectories over time was quantified relative to the origin of the space, defined during the calibration procedure. Movement variability was operationalized as the standard deviation of the distance values from this center, providing an index of the spatial dispersion of the child's motion across each trial. Second, the distance between the child and the caregiver was extracted across the duration of each trial. Distance from the caregiver was quantified by computing the Euclidean distance between the child and the caregiver at each time frame, yielding a continuous estimate of caregiver proximity throughout the trial. As during pre-processing, some of the markers were not visible, parts of the trials had to be removed, resulting in trajectories with different numbers of points. Therefore, the motion trajectories and trajectories of distance over time were then normalized by interpolating it over 100 points, using cubic splines.

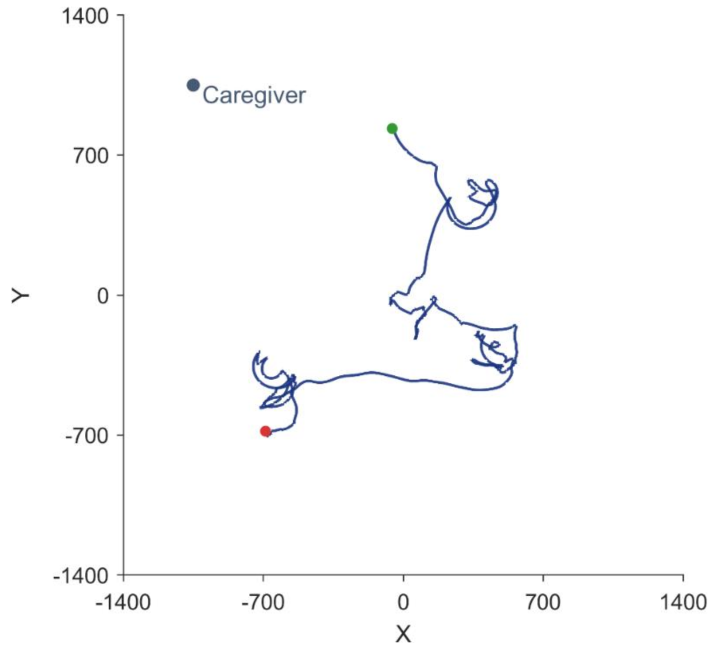

Figure 3.4: Example of a participant's movement trajectory during a single trial. The trajectory starts at the green point and ends at the red point. The grey point represents the caregiver's static position.

3.2.5 Data analysis

Statistical analysis was conducted in Matlab R2023b to evaluate potential differences across conditions in terms of trajectory variability and distances maintained from the parent over the duration of the trials. Median arrays were calculated within each participant for each condition, as the use of the median helps mitigate the influence of excessive variability in the data. The Statistical Parametric Mapping (SPM)1d package (http://www.spm1d.org/) was employed for the analysis of 1D time-series data (Pataky, 2012). Initially, normalized trajectories were assessed for normality using the SPM1d normality test for 1D data. Normality tests reached significance at $\alpha=0.05$, suggesting enough evidence against the null hypothesis of normality. Therefore, a SPM1d parametric one-way repeated measure ANOVA was used ($\alpha=0.05$, iterations: 1000), with Bonferroni corrections, to test for main effects among baseline, multisensory and silent conditions for the two variables of interests. Post hoc comparisons were performed with SPM1d parametric paired t-tests between conditions.

3.3 Results

3.3.1 Findings

Figure 3.5 showed the results regarding the variability of the children's motor trajectory. The main statistics with α=0.05 showed significant effects (SPM (F)= 8.681, p=0.041) among the baseline, multisensory and silent conditions. In particular, post hoc comparisons, with α=0.02, showed a significant lower motion variability with respect to the origin of the space (SPM (t)= 5.138, p=0.012) during baseline compared to the silent condition, approximately between 35-45% of the trial duration. No significant differences were observed between baseline and multisensory as well as between multisensory and silent conditions.

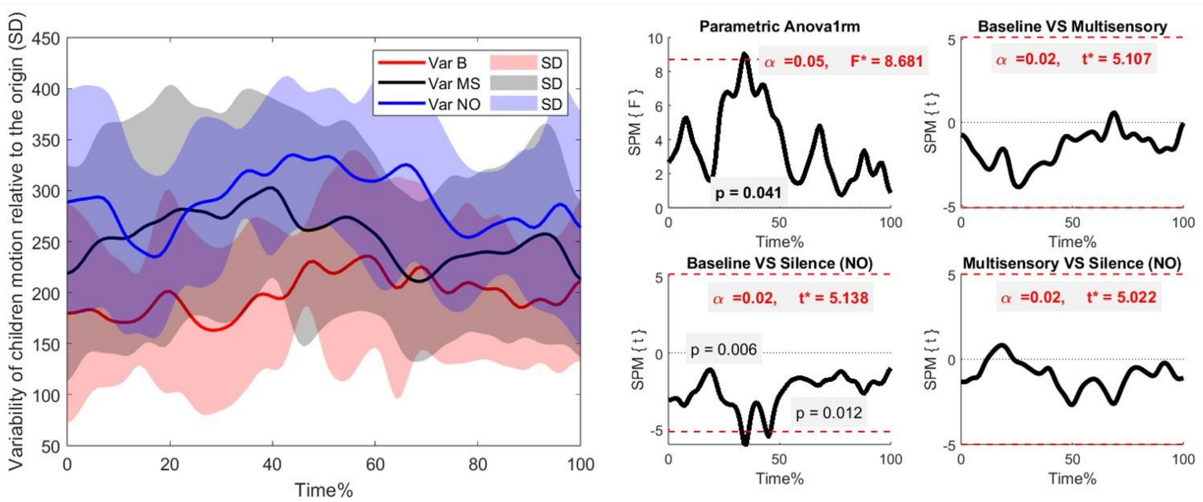

Figure 3.5. Left: Variability of children's motion relative to the space origin over time, for all experimental conditions. The figure demonstrates how the continuous time-series (y-axis) behave over time (x-axis), for all the conditions. B=Baseline, MS=Multisensory and NO=Silent. ***Right: SPM1d parametric Anova1rm*** corrected for multiple comparisons, with SPM1d parametric post hoc paired t-test. The main statistics showed significant main effects. Post hoc comparisons showed statistically significant lower variability of children motion from the origin during baseline compared to silent condition from 35-45% of trial duration.

Figure 3.6 showed the results regarding the median distance between children and parents. The analysis revealed the main statistics with α=0.05 showed significant effects (SPM (F) = 7.512, p=0.006) among the baseline, multisensory and silent conditions. Next, post hoc comparisons with α=0.02 showed statistically significant larger distance kept from the parents during baseline compared to multisensory condition at half duration of the trial (SPM (t)= 4.845, p=0.002). No differences between the baseline and silent conditions, as well as between silent and multisensory have been found.

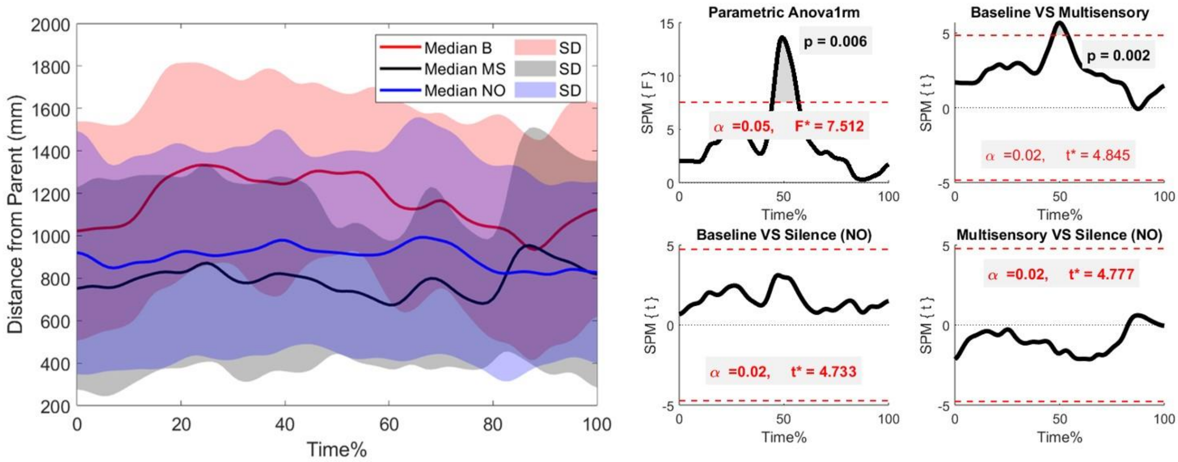

Figure 3.6. Left: Median of distance kept from the parent over time, for all experimental conditions. The figure demonstrates how the continuous time-series (y-axis) behave over time (x-axis), for all the conditions. B=Baseline, MS=Multisensory and NO=Silent. ***Right: SPM1d parametric Anova1rm*** corrected for multiple comparisons, with post hoc SPM1d parametric paired t-test. The main statistics showed significant main effects. Post hoc comparisons showed statistically significant larger distance kept from the parents during the baseline compared to multisensory condition at half duration of the trial.

3.4 Discussion

This study explored whether multisensory cues modulate motor exploration in children with VI and reorganize their interaction with caregivers. When multisensory stimuli were introduced into the playroom, children with VI displayed increased variability in their movement trajectories and remained in closer proximity to caregivers.

A clarification regarding the temporal dimension of the analyses is warranted. In the present study, time was not modeled as a progressive factor across trials. Instead, all kinematic measures were computed within each trial and then averaged across trials belonging to the same experimental condition for each participant. Consequently, the reported effects reflect condition-specific differences averaged at the trial level. Within this framework, movement variability and distance from the caregiver capture different, partially dissociable aspects of behavior. Movement variability indexes the dispersion and stability of the child’s motor activity within a given trial, whereas distance from the caregiver reflects caregiver-directed spatial regulation.

From the analysis children’s motor trajectories were significantly more variable during the silent condition compared to the baseline, which was evident before half the time of trial duration. The multisensory condition also demonstrated larger, albeit not significant, motor trajectory variability compared to the baseline. This means that in both multisensory and silent conditions, less consistent motor behaviors were detected compared to the baseline, as some children might have moved more, and others less. It is important to notice that multisensory and silent conditions were alternated: the silence conditions represented how the child explores the environment when the extraneous stimulus disappears. This significant increase in variability within the first half of trials could be conceived as an attempted motor strategy of children with VI to regain control of their social space after an unexpected stimulus was introduced, as they perceived multisensory cues as novelty. Therefore, there might be different between-children "trial and error" attempts to guide motor activities within an unpredictable environment (Keen, 2011). This could lead to different motor strategies behaviors in relation to having poor visual control over an environment. Vision is, in fact, a crucial sensory channel for the early development of attention (Tadić et al., 2009). Indeed, children with VI show a “developmental setback”: in the absence of integrative visual input, developing attentional control might be more difficult for such children, resulting in consequences in the motor domain (Cass et al., 1994; Dale & Sonksen, 2002). In typical development, the first years of life can be considered as fundamental developmental epoch marked by a growing drive for autonomy and independence. During this period, children’s attention becomes more controlled and focused and is less driven by novelty (Rothbart et al., 2006). It is plausible to state that children with VI perceived the unusual multisensory stimuli as a novelty. Hence, the present study suggests that when children lose control over what they expect to happen within the environment, their motion becomes less consistent.

The findings of this study showed that with multisensory stimuli, children kept a significantly shorter distance from their caregivers compared to the baseline, which was evident at exactly half duration of the trials. Such pattern was apparent also during the silent condition, which however was not significantly different from the baseline at any point in time. As stated above, it is reasonable to state that instead of promoting activity, multisensory stimuli were perceived as unfamiliar triggers. These results can be interpreted with reference to the theoretical framework of the Strange Situation paradigm (Ainsworth et al., 1978), which informed the conceptual development of the experimental protocol. Importantly, the present paradigm did not include parental separation phases nor the physical presence of a stranger; instead, multisensory cues were introduced while the caregiver remained continuously present. Within this constrained framework, the multisensory condition may be understood as introducing environmental novelty or perturbation, rather than as a true substitute for a stranger figure. The observed tendency toward increased caregiver proximity during the multisensory condition may therefore reflect a regulatory response to heightened environmental uncertainty, consistent with attachment-related behaviors. Similarly, the tendency (although not significant) to seek proximity during the silent condition may suggest a form of reunion-like regulation; however, the continuous alternation between multisensory and silent conditions likely limited the temporal window necessary for the emergence of more stable, statistically detectable motor patterns of caregiver-directed behavior. As such, the present findings should be considered preliminary and illustrative of how principles derived from attachment theory may be adapted to study embodied and motor aspects of intersubjectivity in children with VI.

Emotional and motor behavior towards the attachment figure has long been of central interest to developmental psychology: attachment behaviors could be related to problem solving. The balance between attachment and exploration has a central role in shaping children's attentional flexibility (Main, 2000), with emotion and attention regulation being closely linked processes (Pallini et al., 2019). Indeed, high levels of emotional arousal could impair both attentional focus and self-regulatory behaviors (Diamond & Fagundes, 2008). Conversely, attention-regulation strategies are central to managing emotions (Gross, 2013) and contribute significantly to motor development. It is plausible to state that the increased search for proximity to the attachment figure when the unusual stimulus was played, was an attempt to shift the attentional focus to trigger emotional regulation. Indeed, the primary role in attachment lies in proximity seeking, a defining feature of the attachment bond (Bowlby, 1969). The average shorter distance from the attachment figure occurred when the children perceived the multisensory condition as an interference, so that they looked for proximity to their attachment figure at half duration of the trial. This might represent a possible problem-solving strategy related to the perception of an unusual stimulus. In other words, when multisensory information was introduced, there was an attempt from children with VI to continue keeping the same distance from caregivers as in the baseline and silent condition. However, to compensate for the presence of unusual stimulus, proximity to the caregiver was necessary at half trial to then continue with their spontaneous activities.

The present study is in line with previous works sustaining the idea that broadening the focus to include the noisy, rich complexity of natural behavior expands researchers' view (Dahl, 2017; Tamis-LeMonda et al., 2017). Indeed, developmental psychology is not a science of lonely children, although in most studies researchers observe children in isolation. Nevertheless, from an ecological perspective, motor development is a relational enterprise (Adolph & Hoch, 2019). Understanding how limitations in visual information affect executive, motor and then relational functions is essential for designing both relevant research models and effective, needs-based interventions (Gui et al., 2023). Indeed, children with VI encounter distinct difficulties in their relationships with parents and peers, as well as in play and social interactions. Supportive parental caregiving and customized interventions play a vital role in fostering their development**.** Considering the high prevalence of social skills challenges within this population and their impact on all stages of development, it is crucial for practitioners to incorporate social skills assessment and targeted instruction into intervention programs (Caron et al., 2023). In line with that, early interventions targeting families of infants with VI have emphasized the significance of fostering interaction, intersubjectivity, and joint attention to strengthen parent-child relationships (van den Broek et al., 2017a). However, the effectiveness of these interventions depends on their ability to specifically address the aspects of the parent-child relationship most impacted by the VI. The present work provides an interpretation of the motor behavior in relational terms, offering valuable insights into the cognitive and emotional strategies employed by children with VI.

Regarding the use of multisensory stimuli, it is noteworthy that the quality of stimulation (such as the type of sound or color) and the method of presentation (predictable or not) could either promote or hinder motor and social behavior. Interestingly, in a previous study (Verver et al., 2019) audio cues were found to enhance shared attention among children with VI during the exploration of play materials; however, the sensory stimulation appeared to interfere with social interactions during peer play. In line with these previous findings, in the present work the unexpected and discontinuous sensory stimulus generated a variation of motor exploration, with the consequent activation of attachment behavior to the parent. The results obtained are also supported by clinical practice, where it is noted that as a child's visual acuity decreases, their response to new sensory information becomes increasingly inhibited. Consequently, it is the quality and the use of multisensory information over time that determine the effectiveness of a protocol on the child's motor and social behavior.

Despite the novelty of this work, several limitations should be acknowledged. First, the small sample size likely resulted in limited statistical power, reducing sensitivity to detect subtle or moderate effects and constraining the generalizability of the findings. As a consequence, conclusions should be considered exploratory. In addition the absence of a parallel video-coding counterpart and the continuous alternation between multisensory and silent conditions. Indeed, evaluating automated measurements against manual coding is a crucial step in establishing the accuracy and validity of an automated method (Væver et al., 2015). Future work could be redirected, combining video coding procedures with the quantification of motor parameters to compute other significant motor and relational aspects through motion capture systems. Moreover, a longer time of familiarization with the multisensory stimuli would be more ecologically valid, in line with the rehabilitation context and more generally with the real-world environment. Accordingly, the next step of this work will consist in restructuring the conditions and the multisensory stimuli of the present protocol to unveil different motor, attentional and relational strategies of children with VI.

3.5 Conclusion

Embedded intersubjectivity: variability and proximity to caregivers under multisensory uncertainty

By reconceptualizing a protocol inspired by the Strange Situation paradigm to children with VI, the present work revealed how unexpected sensory events modulate kinematic behavior, with proximity acting as a regulator of emotional activation and motor variability emerging as a functional resource for maintaining interaction and environmental control. These findings closely align with the embedded framework: if perception and action are inseparable and situated, and if infants actively generate information through movement, then a reduction in visual input necessarily reorganizes exploratory behavior. In this sense, the increased proximity and movement variability observed in VI children can be interpreted as embodied strategies to sustain regulation and social connection when perceptual information becomes less predictable or less accessible. Aligned with the developmental framework for early intervention (Dale & Salt, 2007) this study proposes that parents can use "forewarning" to prepare children for unexpected and potentially distressing sensory experiences, thereby reducing the risk of sensory aversion. Additionally, caregivers can support the child’s learning across various domains, such as exploring their movement potential, developing spatial awareness and safety mechanisms, and utilizing alternative sensory strategies to understand objects and the environment.

Chapter 4

Enculturated intersubjectivity as otherness.

A multimodal method for detecting joint attention in naturalistic interaction

Summary

This chapter proposes a continuous multimodal framework for detecting joint attention through head orientation, combining clinical coding, computer vision (CV), and 3D motion tracking to capture naturalistic caregiver–child exchanges. It was hypothesized that this approach would complement subjective clinical assessment and that high agreement between marker-based and CV methods would demonstrate the robustness of head orientation as a proxy for identifying joint attention episodes via head orientation. For this, ten mother–child dyads, including children with VI and typically developing peers, participated in five-minute play sessions in an ecologically setting. Head orientation was tracked using forehead markers and estimated using clinical coding, CV and motion capture. A vector-wise approach quantified relational, directional, and disjoint attention by computing angles between caregivers' and children's head orientation vectors. Two computational models Weighted Least Squares (WLS) model and Gaussian Mixture Model (GMM) were tested for robustness of angle classification. The modeled data was then validated by comparing agreement across clinical coding, CV and motion tracking, evaluated using an event-based temporal matching index. Results showed that GMM achieved the highest agreement and was used to identify joint attention episodes. There was also high consistency between CV and motion tracking, while concordance with clinical coding was lower, reflecting temporal misalignment rather than annotation criteria. Regarding joint attention episodes, results revealed similar trends between tools and groups. The proposed framework offers a valid alternative to gaze-based measures, supporting inclusive models of interaction and quantitative assessment of joint attention in populations where gaze metrics are limited, such as children with VI and those with autism spectrum disorder.

4.1 Introduction

4.1.1 Head orientation and joint attention: integrating relational and computational perspective

To demonstrate joint attention, typically developing children alternate their gaze between the adult and the object of interest, not only indicating that they are attending to the same object but also that they are coordinating this shared focus with mutual engagement (Tomasello & Carpenter, 2007). Engaging in joint attention establishes the basis for shared intentionality (Tomasello et al., 2005). In fact, joint attention is the birthplace of infants 'sensitivity to others' mental lives (Moll & Meltzoff, 2011), and it shapes early social cognition, becoming essential to consider the nonverbal mechanisms through which it is enacted. Head orientation, in particular, plays a crucial role in coordinating relational dynamics from the earliest stages of development: infants use head orientation to regulate spatial boundaries and the degree of engagement in face-to-face interactions (Beebe & Stern, 1977; Brazelton et al., 1974). Since head and eye movements often operate in tandem during interactions, they jointly signal attention and intention (Hanna et al., 2020). While eye gaze is normally a more precise and reliable indicator of attention compared to head orientation, the latter still offers meaningful information, especially regarding the direction of focus (Langton & Bruce, 2000). Moreover, head orientation serves as a larger visual cue, which is more easily detectable in peripheral vision than eye gaze (Loomis et al., 2008). Previous works have shown that head orientation accounts for an average of about 70% of gaze direction (Stiefelhagen & Zhu, 2002) and allows for accurate identification of the target being observed in 89% of cases (Waibel et al., 2003). Building on this findings, researchers have developed increasingly efficient computer vision algorithms for estimating head orientation (Murphy-Chutorian & Trivedi, 2008). These advancements reflect a growing understanding of the communicative significance of head poses estimation within the broader field of human pose analysis (Gao et al., 2025). Such technological advances not only improve precision in tracking but also open new possibilities for understanding the social functions of head movement (Morales et al., 2000). Building on this, time series analysis within a naturalistic mother-infant setting revealed that variations in head angular amplitude corresponded with moments of communicative mismatch (Væver et al., 2015). These findings suggest that head movements can offer valuable insights into emotional coordination in primary interpersonal exchanges (Hammal & Cohn, 2014; Hammal et al., 2013). In typical dyadic exchanges, variations in head orientation help modulate contact with the caregivers, who in turn adjusts their own positioning in response (Lavelli & Fogel, 2005; Weinberg & Tronick, 1994). Beyond typical development, head movement analysis has also shown promises in identifying atypical social patterns, particularly in children with Autism Spectrum Disorder (ASD). Emerging evidence suggests that even spontaneous head movements in early childhood might carry diagnostic value (Gima et al., 2018) and metrics derived from head movement can be used to distinguish children with ASD from TD peers (Zhao et al., 2022). Another previous experimental protocol involving children with ASD used a humanoid robot to elicit behavioral metrics from body and head movements during a task for the assessment of joint attention (Anzalone, 2019). These findings support the diagnostic and developmental relevance of head orientation in populations with atypical social trajectories. While much of the existing research has focused on children with TD and ASD, little is known about how these nonverbal mechanisms unfold in children with VI, who face distinct challenges in achieving coordinated joint engagement. Congenital VI has been linked to significant delays across multiple developmental domains, particularly in social interaction (Tadić et al., 2009).

Accordingly, these infants may not perceive their caregivers' visual focus or interest (Dale & Salt, 2007) and consequently may have difficulties in engaging in contingent and mutually responsive exchanges with them (Bigelow, 2003). According to a recent review (Grumi et al., 2021), certain postural or face signals observed in infants with VI are often interpreted by caregivers as a lack of interest, reducing positive social interactions. These behaviors include limited or absent responses to eye contact, gaze, or smiling during social interactions (Lueck, 2008), turning their gaze and head away from caregivers (Alfaro et al., 2021) and showing difficulties engaging in joint attention with peers (Verver et al., 2020).

Despite its critical importance, research on development of joint attention in infants and children with VI remains scarce. Furthermore, traditional approaches for assessing joint attention episodes (e.g., measuring an infant's gaze-switching behavior between a social partner and an object), are less applicable to pediatric populations with VI (Alfaro et al., 2018; Alfaro et al., 2021). Since research on joint attention has predominantly focused on the visual modality (Akhtar, 2007), alternative approaches suitable for infants and children with VI remain underexplored. This gap is particularly striking given the limitations of gaze-based measures in children with VI and the need for alternative indicators of joint attention and relational engagement. Therefore, the aim of this study was to introduce and validate a multi-modal, vector-based framework for identifying joint attention episodes from head orientation during caregiver–child interactions, in both children with TD and VI. It was hypothesized that the proposed approach would be robust across tracking techniques (clinical video coding, computer vision and 3D motion tracking), and would complement qualitative clinical assessments by providing continuous, quantifiable metrics of interaction.

4.2 Methods

4.2.1. Participants and experimental set up

This study has been carried out as part of collaboration between the Center of Child Neuro-Ophthalmology at IRCCS Mondino Foundation (Pavia, Italy), and the Unit for Visually Impaired People at Istituto Italiano di Tecnologia, (Genoa, Italy). The experimental protocol was approved by the local Ethics Committee (Fondazione IRCCS Policlinico San Matteo, Pavia, Italy) and parents gave their written informed consent under the Declaration of Helsinki.

The inclusion criteria were: a) VI of peripheral origin (with an involvement of the pregeniculate visual pathway); b) visual acuity < 0.5 logmar scale, and c) the ability of walking confidently for at least one year i.e., being able to take more than 20 steps without falling or losing balance, pick up toys while walking, navigating around objects and interact with caregivers while walking (Price et al., 2022). Participants were recruited from the clinical population in care at the Centre of Child Neurophthalmology, IRCCS Mondino Foundation, Pavia (Italy). According to the criteria described above, five children with Typical Development (TD) and five children with VI were selected to take part in the present study. The target age from 2 to 4 years of age was chosen because, in children with VI, the most serious consequences for development are in those children showing a developmental setback in their second to third year, with plateauing or loss of cognitive and language skills and increasing disorder of social communication skills (Dale & Sonksen, 2002). The experimental procedure was explained to the parents of the participants prior to testing. During the session, dyads were encouraged to interact as naturally as possible, without specific instructions, with the only constraint that the caregiver remained seated on a chair positioned in one corner of the setup. The set-up was the same as a previous work: an ecological playground with a dimension of 3x3 m with a gate of 1 m of height (Guarischi et al., 2025).

The following section describes the different approaches adopted to assess attentional states. First, clinical video coding is presented: this approach did not involve the construction of vectors but relied on the clinician's direct categorization of three predefined attentional states based on observational criteria. The section then introduces the computational approaches based on computer vision and 3D motion capture, in which two vectors, one for the caregiver and one for the child, were constructed (Figure 4.1). These vectors were used to compute the interactive angle between the two agents, which served as the basis for identifying three types of attentional states.

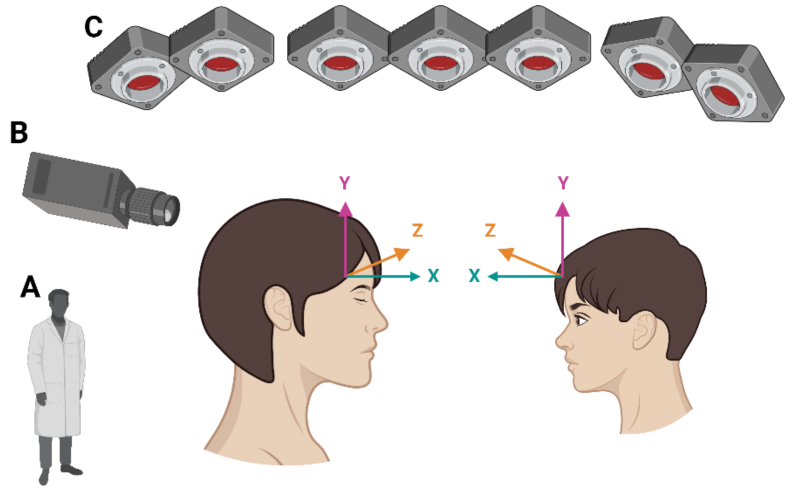

Figure 4.1: Coding techniques. A. Clinical video coding B. Computer vision C. 3D motion capture

4.2.2 Clinical video coding

Instrumentation specifics and data pre-processing

Data acquisition was carried out using an RGB video camera (Canon LEGRIA HF R88). The camera was positioned in a frontal position at 1,5 m of distance from the gate of the setup. CapCut (ByteDance, Version 7.0.0) was used to trim the video in order to synchronize it with the motion capture trial, ensuring that both recordings shared the same starting and ending phases. A clinician then coded the video using Behavioral Observation Research Interactive Software (BORIS) (Friard & Gamba, 2016) with a temporal resolution of one second. The instructions were to observe each frame and classify the head orientation between the dyads as one of three categories: relational (heads facing each other), directional (heads pointing in the same direction), or disjoint (heads pointing in opposite directions).

4.2.3 Computer vision

Instrumentation specifics and data pre-processing

As for section 2.2, data acquisition was carried out using an RGB video camera (Canon LEGRIA HF R88). Then, computer Vision and Machine Learning algorithms to detect head orientation from videos were adopted.

Head pose estimation and attentional vector construction

In parallel with the motion capture system, head orientation was estimated using a computer vision pipeline based on RGB video input. The method relies on HHP-Net (Tomenotti et al., 2024), a lightweight neural network architecture capable of predicting head pose angles as triplets of (yaw, pitch, roll) angles with associated uncertainty. The triplets of angles with their uncertainties are estimated with an end-to-end architecture using the output of a pose detector as input. In particular, the keypoints located on the faces of the subjects in the scene are provided to the network. Head pose detection is applied frame-by-frame and then used to construct attentional vectors for each agent.

The input of HHP-Net is composed of 5 keypoints located on the face (eyes, nose, ears), obtained by a pose detector (Cao et al., 2019). Head Pose Estimation is then shaped as a regression problem, to simultaneously learn end-to-end 6 outputs, i.e. the 3 angles with their uncertainty. For the latter, the method considers the aleatoric heteroscedastic uncertainty, i.e. uncertainty due to input data variability, and it represents a cue for network interpretation.

To visualize the Head Pose in the image plane, it is convenient to project the yaw-pich-roll (q_y, q_p, q_r) triplet using the Tait-Bryan angles and assuming the application is the nose 2D point $X = (x, y)$ in the image. Of particular relevance for our purposes is the vector with ending point $X_h = (x_h, y_h)$ defined as

$$x_h = \sin q_y + x$$

$$y_h = -\cos q_y * \sin q_p + y$$

Given a video sequence, HHP-Net runs in each frame separately, providing a head pose for each person visible in the scene. Hence, at a certain time t, there are two interesting head positions, the one of the mother X^M and the one of the child X^C, and the corresponding head poses X_h^M and X_h^C.

4.2.4 Motion capture

Instrumentation specifics and data pre-processing

During ecological interactions, 3D motion data was tracked with the Vicon System (Nexus, Vicon Motion Systems Ltd. UK) using eight near-infrared cameras (Vero cameras, 100 Hz). Spatial calibration was performed in the Vicon Nexus Software before data capture started, according to the system specifications. Then, a 3D model was created using the head as a single kinematic segment where three markers were identified (left, central and right foreheads), for each agent, i.e. both parents and children, who wore head bands with three markers that were placed on the right and left temples, as well as on the middle forehead.

Following data pre-processing in Nexus Software, the x, y and z coordinates of the three forehead markers were extracted from the .c3d files from a custom written Matlab script (v R2023b, The MathWorks, Natick, USA). These coordinates were then filtered using a low-pass Butterworth filter with a 6 Hz cutoff to remove any residual noise in the data (Winter et al., 1974). For the present study, the motion trajectories of all the three markers of both children and parents were utilized.

Local reference system construction and vectors construction

In order to track head orientation of caregivers and children in the 3D capture space, a local reference system was needed for each agent. This meant creating a frame of reference that was expected to be integral with the head of each agent. By definition, a frame of reference is formed by three unitary and perpendicular (among each other) vectors. Here, we reported the mathematical preliminaries of the local reference system creation for each agent. Firstly, the x axis is created: one for the mother ($\vec{\boldsymbol{x}}_{\boldsymbol{m}}$) and one for the child ($\vec{\boldsymbol{x}}_{\boldsymbol{c}}$).

$$\vec{\boldsymbol{x}}_m = \frac{\overline{m_1}_{|SG} - \overline{m_2}_{|SG}}{\left\| \overline{m_1}_{|SG} - \overline{m_2}_{|SG} \right\|} \quad (1) \qquad \vec{\boldsymbol{x}}_c = \frac{\bar{c_1}_{|SG} - \bar{c_2}_{|SG}}{\left\| \bar{c}_{|SG} - \bar{c}_{|SG} \right\|} \quad (2)$$

In equation 1 $\overline{\boldsymbol{m_1}}_{|\boldsymbol{SG}}$ indicate the front marker of the mother and $\overline{\boldsymbol{m_2}}_{|\boldsymbol{SG}}$ refers to a marker constructed as the average position calculated between the right and the left markers. The letters **SG** indicates that the markers coordinates are referred to the global reference system, i.e. the motion capture coordinate system set during the spatial calibration.

In equation 2 $\overline{\boldsymbol{c_1}}_{|\boldsymbol{SG}}$ indicate the front marker of the child and $\overline{\boldsymbol{c_2}}_{|\boldsymbol{SG}}$ refers to back marker of the child. As for the mother, this last one was calculated as mean value between the right and the left markers.

$$\vec{\boldsymbol{t}} = \frac{\overline{m_3}_{|SG} - \overline{m_2}_{|SG}}{\left\| \overline{m_3}_{|SG} - \overline{m_2}_{|SG} \right\|} \quad (3) \qquad \vec{\boldsymbol{t}} = \frac{\bar{c_3}_{|SG} - \bar{c_2}_{|SG}}{\left\| \bar{c_3}_{|SG} - \bar{c_2}_{|SG} \right\|} \quad (4)$$

In equation 3, $\overline{\boldsymbol{m_3}}_{|\boldsymbol{SG}}$ indicate the marker of the right and $\overline{\boldsymbol{m_2}}_{|\boldsymbol{SG}}$ is the marker as defined earlier.

In equation 4, $\overline{\boldsymbol{c}}_{|\boldsymbol{SG}}$ indicate the marker of the right and $\overline{\boldsymbol{c_2}}_{|\boldsymbol{SG}}$ is the marker constructed on the back of the children's head.

$$\vec{\boldsymbol{z}}_m = \vec{\boldsymbol{t}} \ x \ \vec{\boldsymbol{x}} \quad (5) \qquad \vec{\boldsymbol{z}}_c = \vec{\boldsymbol{t}} \ x \ \vec{\boldsymbol{x}} \quad (7)$$

$$\vec{\boldsymbol{y}}_m = \vec{\boldsymbol{z}} \ x \ \vec{\boldsymbol{x}} \quad (6) \qquad \vec{\boldsymbol{y}}_c = \vec{\boldsymbol{z}} \ x \ \vec{\boldsymbol{x}} \quad (8)$$

In equation 5, the z coordinates of the mother vector were calculated as a product of t and x vectors.

In equation 6, the y coordinates of the mother vector were calculated as a product of z and x vectors. In equation 7, the z coordinates of the child vector were calculated as a product of t and x vectors. In equation 8, the z coordinates of the child vector were calculated as a product of z and x vectors. The resulting coordinate axes for the child ($\vec{\boldsymbol{x}}_{\boldsymbol{c}}, \vec{\boldsymbol{y}}_{\boldsymbol{c}}, \vec{\boldsymbol{z}}_{\boldsymbol{c}}$) and the mother ($\vec{\boldsymbol{x}}_{\boldsymbol{m}}, \vec{\boldsymbol{y}}_{\boldsymbol{m}}, \vec{\boldsymbol{z}}_{\boldsymbol{m}}$) represent the new local reference systems.

4.3 Defining joint attention episodes: classification of relational, directional and disjoint attention

Once the computation of the local reference system was developed using the 2D computer vision and 3D motion capture tracking techniques, for both caregivers and children, the dyadic interactions were classified into three categories: relational, directional and disjoint, aiming to identify episodes of joint attentions. To do so, a cut-off angle selection was assigned to each attentional category. Then, this logic was applied to the angles forming between the vectors constructed for caregivers and children using the local reference systems. Below the passages leading to angle selection were described.

4.3.1 Angles identification and calculation

To define angles ranges for classification of the three attention categories mentioned above, a motion capture system was used as the gold standard. Then, the following procedure was adopted: two individuals sat in front of each other, one individual remained in a fixed position while the other rotated by 360°. The moving individual used an electronic goniometer to capture their degree of motion precisely every 30°. Those positions were then used to calculate the angles forming between the vectors originating from the local reference system of caregivers and children heads. Specifically, the x axis of each local reference system represents the 3D head-direction vector (i.e., the forward pointing direction of the head). Accordingly, δ is defined as the 3D angle between the head-direction vectors of mother and child.

$$\delta = atan2(\|\vec{\boldsymbol{x}}_{\boldsymbol{c}} \times \vec{\boldsymbol{x}}_{\boldsymbol{m}}\|, \vec{\boldsymbol{x}}_{c} \cdot \vec{\boldsymbol{x}}_{m})$$

Where $\vec{x}_c \times \vec{x}_m$ is the cross product between the two vectors, creating a resulting perpendicular vector; $\vec{x}_c \cdot \vec{x}_m$ is the scalar product, referring to how aligned the two vectors are (whether parallel or not) and $atan2$ is the function returning the angles. The unsigned angle between the two vectors was computed. Since atan2 returns values in the range 180° -180°, using the magnitude of the cross product in the numerator ensures that the resulting angle is always between 0° and 180°.

4.3.2 Attentional categories cut-offs

Relational attention is defined as the condition when both agents are looking at each other Based on the angle identification, this interaction corresponded to angles between 150° and 180° and were coded as –2. Directional attention describes situations where the agents are oriented in the same direction, which angles were defined ranging from 0° to 30°, and coded as 0. Disjoint attention refers to instances in which the agents are looking in different directions but not toward each other, falling within the range of 30° to 150°, which were coded as 1.

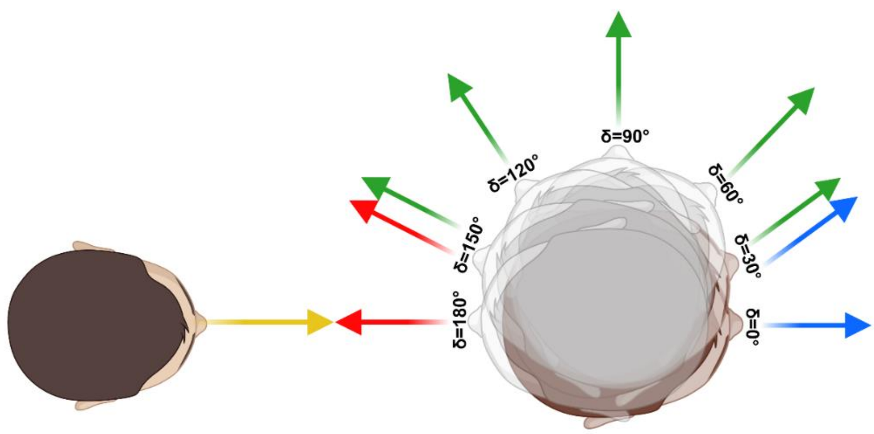

Figure 4.2: Representation of attentional category. The displayed angle ranges from 0° to 180°. Classification of attentional category: directional (0°-30°, blue arrows), disjoint (30° -150°, green arrows) and relational (150°-180°, red arrows).

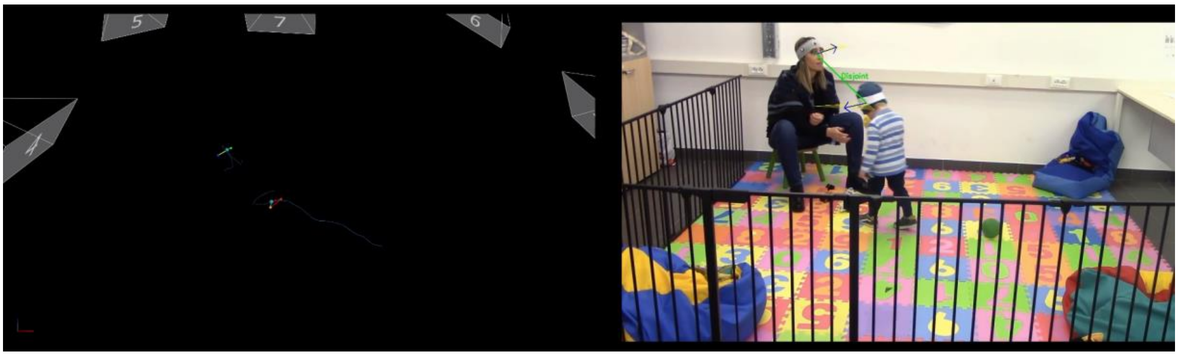

Figure 4.3: Example of the same frame for 3D motion capture data (left) and for computer vision data (right).

Before applying predefined cutoff ranges for attentional categories, a preliminary check was conducted on a limited set of frames to ensure that the selected angular thresholds corresponded to head orientations consistent with the intended attentional categories. This verification involved comparing angles calculated from motion capture data with visual inspection of video recordings, confirming that orientations near the defined boundaries reflected the expected attention configurations. While this categorical approach provides clear interpretability, it relies on rigid thresholds that may not fully capture the graded nature of attentional dynamics. To address this limitation, two complementary models were developed: a WLSM, which integrates angular relations with regression-based weighting, and a GMM, which leverages probabilistic clustering to capture variability in head orientation strategies. The following sections describe these models in detail.

4.3.3 Probabilistic models for angle classification

Two computational models were introduced to improve the distinction between attentional states beyond rigid angular thresholds: a Weighted Least Squares (WLS) model and a Gaussian Mixture Model (GMM).

Weighted Least Square (WLS) model

The WLS model defines three Gaussian functions a priori, each centered on a semantically meaningful angular configuration. The three Gaussian functions served as weighting profiles, assigning higher influence to samples whose interaction angles closely matched the attentional prototypes: 0° for directional attention**,** 90° for disjoint orientation**,** and 180° for relational alignment (Fig. 4.4.(A)). To operationalize this, each interaction angle (δ) was represented as the sum of three Gaussian similarity functions, computed as:

$$g_{ij} = \exp\left(-\frac{(\delta_i - c_j)^2}{2\sigma^2}\right)$$

where $c_j \in \{0°, 90°, 180°\}$ are the canonical centers. The standard deviation (σ) determines the spread of each Gaussian weighting profile, influencing the degree to which proximity to canonical centers affects weighting. In this model, σ was fixed at 15°, a value consistent with tolerance thresholds previously adopted in head pose estimation benchmarks, where angular deviations within ±15° are considered acceptable (Ruiz et al., 2018).

For each sample i, a fuzzy target value was derived as the weighted average of these centers:

$$Target_i = \sum c_j * g_{ij} \; / \sum g_{ij}$$

Weights (w_i) emphasize samples near key spatial configurations, capturing the graded nature of attentional states.

$$w_i = \sum g_{ij}$$

A weighted linear regression model was then fitted using the δ angle as predictor and the fuzzy target value as the dependent variable, ensuring that samples closer to canonical configurations exerted greater influence on the fit. The trained model was subsequently used to generate continuous predictions for each observed angle, providing a smooth representation of attentional states. Finally, these continuous outputs were discretized into three attentional categories using predefined thresholds (0°–30° for directional, 30°–150° for disjoint, and 150°–180° for relational). This Gaussian-weighted strategy combines the interpretability of categorical coding with the flexibility of continuous modeling, allowing smooth transitions between attentional states (Fig. 4.4. (B)).

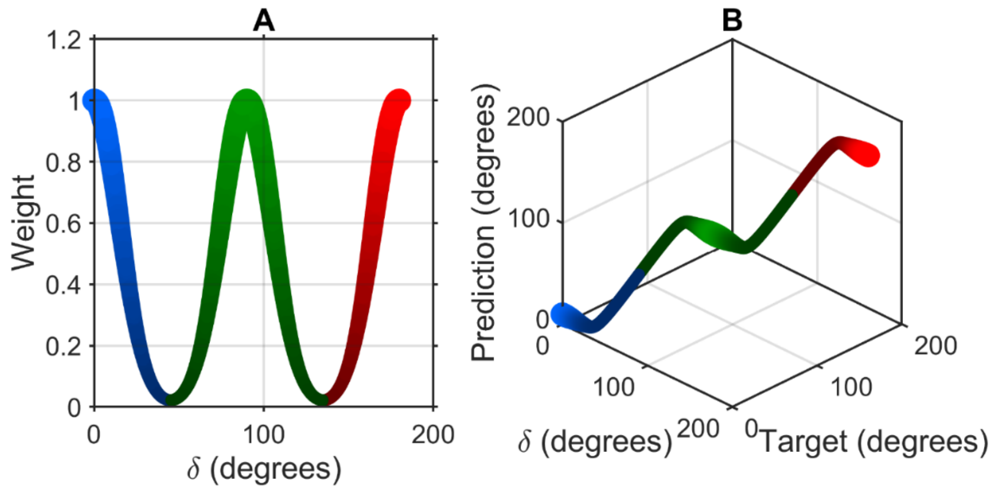

Figure 4.4: **(A) Weight distribution by δ**: Gaussian weighting profiles used to model attentional states based on interaction angles between mother and child. Each curve represents a Gaussian

function centered on a semantically meaningful configuration: 0° for directional attention (blue), 90° for disjoint orientation (green), and 180° for relational alignment (red). The size of each point is proportional to its assigned weight, visually emphasizing the relative influence of each sample in the regression model. **(B) Attentional categorization with weighted regression**: 3D scatter plot illustrates the relationship between the δ angle, the Target (weighted average of centers), and the prediction derived from the weighted linear regression model. Each point represents a time sample, with its color indicating the dominant Gaussian component (centered at 0°, 90°, or 180°) used for weighting, and its size proportional to the computed weight.

Gaussian Mixture Model (GMM)

In contrast to the previous method, which relied on predefined angular cutoff, this second approach adopts a fully data-driven perspective, allowing attentional states to emerge from the empirical distribution of interaction angle (δ) rather than imposing rigid thresholds. The interaction angle was modeled using a GMM, which is well-suited for representing multimodal distributions, with three components estimated via maximum likelihood. The number of components was chosen to align with the three attentional categories defined in the theoretical framework, ensuring semantic interpretability. To prevent singular covariance matrices, a known limitation of Expectation-Maximization (EM) for Gaussian mixtures (Fraley & Raftery, 2002), a small regularization term (1×10^{-5}) was added to diagonal entries to ensure numerical stability. To mitigate sensitivity to initialization, the multiple restart method was adopted (You et al., 2023), which involves running the EM algorithm several times with different random initial conditions and selecting the solution with the highest log-likelihood. Then, the maximum iteration limit was set to ensure convergence while preventing excessive computation time: EM typically stops either when parameter changes are negligible or when a predefined iteration cap is reached (Bonald, 2019).

The mixture model assumes that the observed distribution can be expressed as:

$$\boldsymbol{p(\delta) = \sum_{k=1}^{3} \pi_k \, \mathcal{N}(\delta \mid \mu_k, \sigma_k^2)}$$

where each component is characterized by its mean (μ_k), variance (σ_k^2), and weight (π_k). After fitting the model, the components were ordered by ascending mean angle and mapped to the three attentional categories: directional (low δ), disjoint (mid δ), and relational (high δ). This mapping preserves semantic interpretability while remaining grounded in the data. Each sample was then assigned to the component with the highest posterior probability, producing a categorical label for every observation. By inferring clusters directly from the observed distribution, this approach captures the natural structure of attentional states without relying on arbitrary angular cutoffs, offering a flexible and empirically grounded representation of interaction patterns (see Fig. 4.5 (A) for motion capture data (B) for computer vision data).

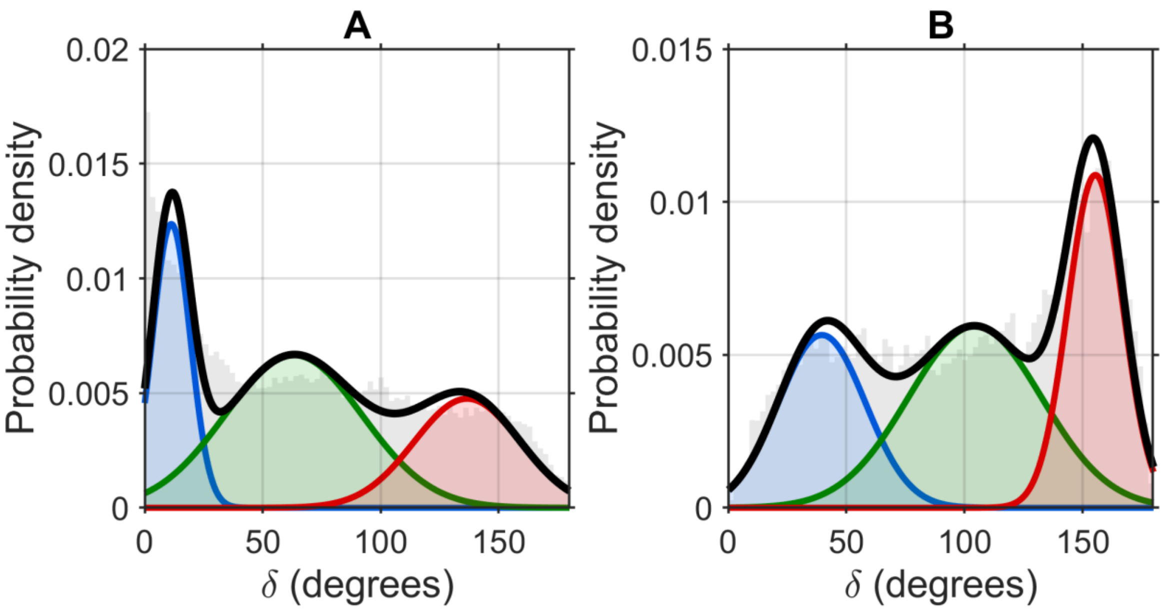

Figure 4.5.: Global distribution of interaction angle (δ) computed from motion-capture data (A) and computer vision data (B) overlaid with the three Gaussian components estimated by the mixture model. Each curve represents one attentional cluster, Directional (blue), Disjoint (green), and Relational (red), with the dashed black line indicating the overall mixture density.

4.4 Statistical analysis

Statistical analyses were carried out using R (version 4.2.2). Agreement patterns were examined across the cutoff-based, WLS, and GMM datasets, and Chi-square tests were performed to assess group differences in the distribution of attentional states obtained from clinical video coding and motion-capture cutoff datasets. The detection of joint attention episodes based on computer vision and 3D motion-capture data was conducted in MATLAB R2022b.

4.4.1 Agreement of classification model using clinical video coding, computer vision and 3D motion capture

To validate the proposed framework, temporal resolution was standardized across modalities. CV data (30fps) was aggregated into bins of 30 frames, and motion capture data (sampled at 100 Hz) were aggregated into bins of 100 samples, matching the one-second resolution of the clinical video coding. For each bin, the most frequent attentional category (mode) was computed. Once this standardization was created timewise, agreement was then evaluated in terms of classification of attentional categories derived from head orientation angles. This evaluation considered two levels of comparison: the consistency between the two automated methods (computer vision and 3D motion capture) and the consistency between each automated method and clinical video coding. The procedure was applied to both datasets generated by the classification models (WLS and GMM), with the aim of determining which dataset provided the most reliable alignment for subsequent analyses of joint attention episodes. Agreement was computed at the event level using an event-based temporal matching index, with a behavior-specific tolerance window defined as ±1 SD of the timing for each behavior. This adaptive window accounts for natural variability in behavioral timing and allows for flexible matching between annotations that are conceptually aligned but not perfectly synchronized in time (Bakeman & Quera, 2011). Regarding interpretation, values closer to 1 indicate higher temporal alignment between annotations.

Clinical video coding relies on discrete categorical annotations rather than continuous head orientation trajectories, so computational models could not be directly applied to this dataset. To assess consistency between groups and determine whether differences in agreement were attributable to temporal misalignment rather than annotation criteria, chi-square tests with Bonferroni-adjusted post-hoc comparisons were conducted on the distribution of attentional categories (relational, directional, disjoint) across TD and VI groups. Clinical dataset was compared with the rigid cutoff-based motion capture dataset, considered the gold standard for consistency in the distribution of attentional categories.

4.4.2 Detection of joint attention episodes using clinical video coding, computer vision and 3D motion capture

The final phase of the present work consists in a temporal pattern analysis for detecting joint attention episodes. This approach is grounded in the idea that joint attention involves not only simultaneous engagement with the same object or event, but also a shared awareness of that engagement (Tomasello, 2014). In line with this view, joint attention is conceptualized here as a dynamic sequence involving the alternation between relational and directional attention, rather than a discrete moment. Drawing on the methodology adopted in previous work (Abney et al., 2017), a minimum duration criterion was established to identify meaningful attentional episodes. Unlike eye-tracking data, which captures rapid and fine-grained shifts in gaze direction, head orientation reflects broader and slower motor patterns (Fang et al., 2015). Therefore, a temporal threshold of 2 seconds was adopted, below which attentional shifts were considered brief transitional states. This choice ensures that detected episodes reflect ecologically valid interactional dynamics, given the slower nature of head orientation signals. Based on the coding conducted for method validation, specific patterns were examined, particularly the sequence of minimum 2 seconds of relational attention followed by minimum 2 seconds of directional attention. Additionally, a parameter called temporal tolerance was introduced to define joint attention episodes based on head orientation. This parameter specifies the maximum number of consecutive non-valid frames that can be tolerated within a joint attention episode, allowing for a more flexible modeling of interactive behavior. Conceptually, it is comparable to the minimum duration used in the eye-tracking study (Abney et al., 2017) which allows for brief deviations in gaze without interrupting a joint attention episode. Similarly, temporal tolerance accommodates short interruptions in head orientation, enabling a more ecologically valid representation of interactional dynamics. The effect of different levels of temporal tolerance (1, 2, 3, 4 seconds) was qualitatively examined across the three tools.

4.5 Results

4.5.1 Agreement of classification model using clinical video coding, computer vision and 3D motion capture

The trend in agreement was consistent across cutoff, WLS, and GMM datasets. Agreement between the angles captured with computer vision and 3D motion capture was high (0.88; 0.94), indicating strong consistency in angle-based classification across motion capture and computer vision. In contrast, agreement between automated and clinical coding was lower (0.29) for both datasets. (Tables 4.1 and 4.2). GMM dataset achieved the highest agreement between computer vision and motion capture and for this reason, it was used in the following analysis to identify episodes of joint attention.

	Video coding vs 3D motion capture	Video coding VS computer vision	3D motion capture VS computer vision
Event-based index	0.29	0.29	0.88

Table. 4.1: Agreement between clinical coding, 3D motion capture and computer vision tools, using WLS dataset.

	Video coding vs 3D motion capture	Video coding VS computer vision	3D motion capture VS computer vision
Event-based index	0.29	0.29	0.94

Table. 4.2: Agreement between clinical coding, motion capture and computer vision tools, using GMM dataset.

Chi-square tests were used to examine group differences (TD vs. VI) in the frequency of attentional states (relational, directional, disjoint) across clinical video coding and motion capture cut off datasets. For clinical coding, the overall chi-square test was highly significant, $\chi^2(2) = 108.92$, $p < .001$, with Bonferroni-adjusted post-hoc comparisons revealing significant differences for all categories: Directional ($p < .001$), Disjoint ($p < .001$), and Relational ($p < .001$). For motion capture data, the chi-square test was also significant, $\chi^2(2) = 20.22$, $p < .001$, with post-hoc comparisons confirming differences for Directional ($p = .0196$), Disjoint ($p < .001$), and Relational ($p = .0089$).

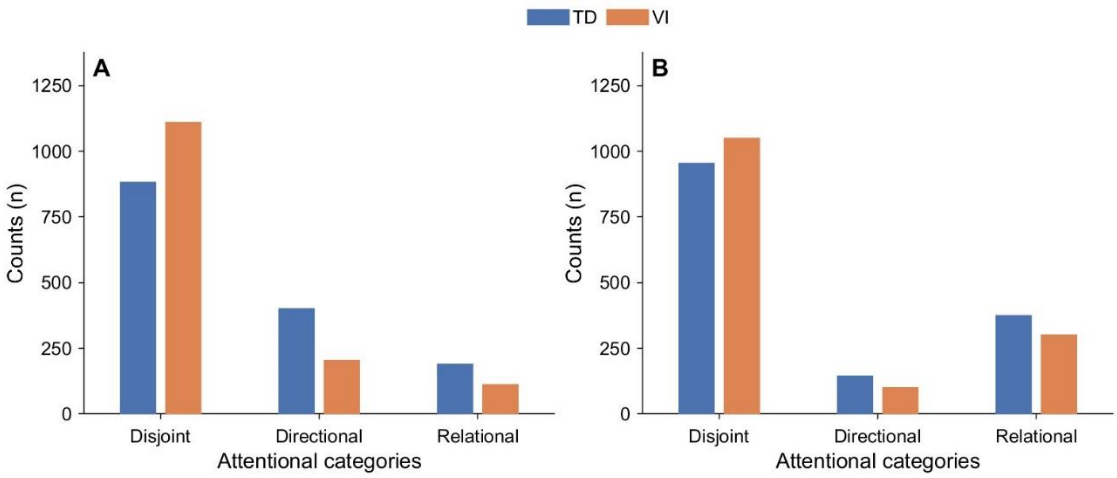

Figure 4.6. Grouped bar chart showing the counts of attentional categories (Directional, Disjoint and Relational) for TD and VI children with clinical coding (A) and 3D motion capture (B) dataset. The chart highlights differences in how often each group displayed the three attentional states, with taller bars indicating higher frequencies.

4.5.2 Detection of joint attention episodes using clinical video coding, computer vision and 3D motion capture

The visual inspection of trial-level plots highlights a strong coherence between CV and motion capture outputs, not only in the classification of attentional categories but, more importantly, in the identification of joint attention episodes. In both modalities, episodes appear as contiguous segments where relational and directional states occur in sequence, and these regions are consistently marked across trials. The highlighted yellow bands indicate the temporal windows classified as joint attention, showing substantial overlap between the two automated methods (Fig. 4.7).

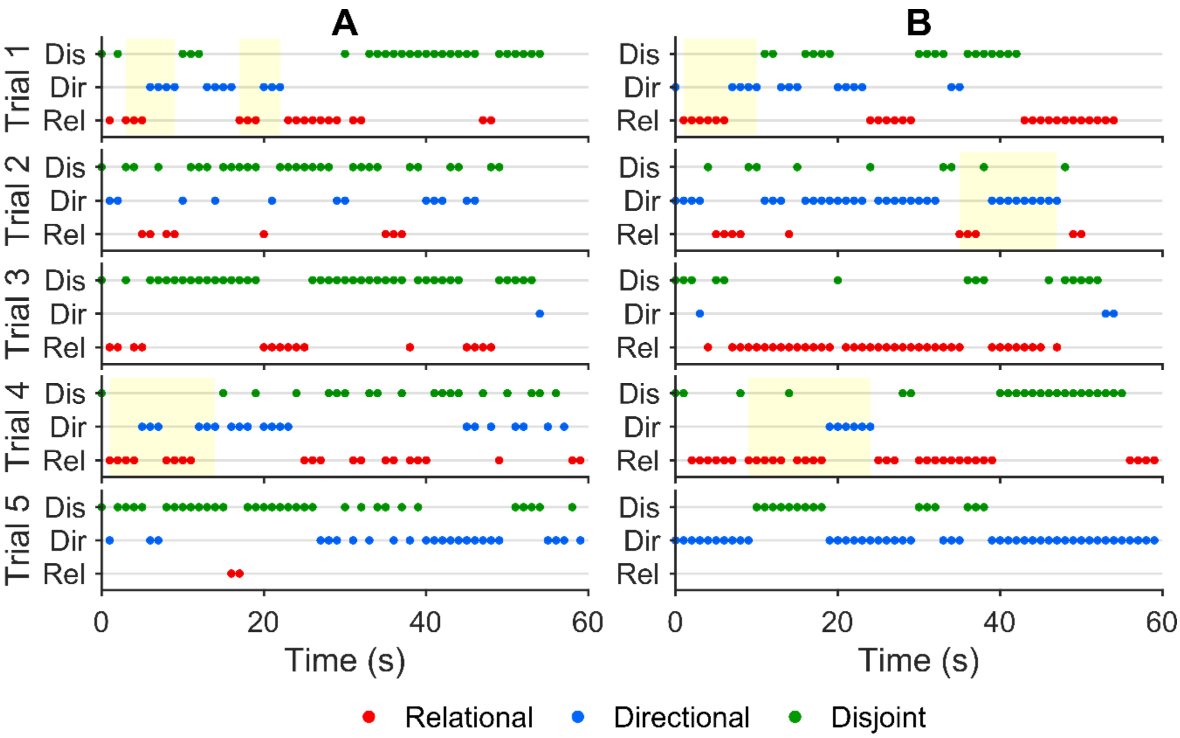

Figure 4.7: Example from one participant showing detected episodes of joint attention with CV (A) and 3D motion capture (B) coding. Color coding follows the scheme described earlier: red indicates relational attention (−2), blue indicates directional attention (0), and green indicates disjoint attention (1). Joint attention episodes are highlighted in yellow.

By using an increasing temporal tolerance, a systematic increase in the number of detected episodes of joint attention was observed. Across all tolerance parameters, a consistent trend emerged between episodes identified via 3D motion capture and computer vision. Moreover, based on a purely descriptive comparison, the overall count of joint attention episodes was comparable across the two groups of dyads: children with VI exhibited a similar number of episodes to their TD peers (Table 4.3 and 4.4).

	Temporal tolerance 1 (s)	Temporal tolerance 2 (s)	Temporal tolerance 3 (s)	Temporal tolerance 4 (s)
TD	18	22	24	27
VI	21	22	25	25

Table 4.3. Episodes count of joint attention episodes across varying temporal tolerance parameters, comparing counts from GMM dataset based on 3D motion capture.

	Temporal tolerance 1 (s)	Temporal tolerance 2 (s)	Temporal tolerance 3 (s)	Temporal tolerance 4 (s)
TD	18	20	22	25
VI	19	20	23	23

Table 4.4 Episodes of joint attention across varying temporal tolerance parameters, comparing counts from GMM dataset based on computer vision.

4.6 Discussion

The goal of this study was to develop a method for detecting joint attention episodes during ecological mother–child interactions using clinical video coding, CV and marker-based motion tracking. It was hypothesized that this approach would complement subjective clinical assessment and that high agreement between marker-based and CV methods would demonstrate the robustness of head orientation as a proxy for identifying joint attention episodes via head orientation. Results showed that GMM achieved the highest agreement and was used to identify joint attention episodes. There was also high consistency between CV and motion tracking, while concordance with clinical coding was lower, reflecting temporal misalignment rather than annotation criteria. The reduced concordance with clinical ratings should be interpreted in light of the different observational resolutions and analytic aims of the respective methods. Clinical coding relies on integrative, retrospective judgments, which makes it less suited to capture brief, temporally precise attentional transitions. In contrast, the automated motion-based and computer vision approaches operate at a fine temporal scale and apply consistent, frame-by-frame criteria, enabling a more reproducible identification of joint attention episodes as dynamically unfolding events. From this perspective, differences between clinical and automated methods reflect complementary levels of analysis rather than limitations of a single approach. Regarding joint attention episodes, results revealed similar trends between automated tools, indicating robust convergence across methods and supporting the reliability of the framework in capturing joint attention dynamics.

Given the relatively small sample size, the present findings should be interpreted primarily from a methodological perspective. The study was designed to evaluate convergence and consistency across measurement tools, rather than to test population-level hypotheses. As such, statistical power to detect subtle effects is limited, and the value of the present work lies in establishing the feasibility and robustness of a multimodal framework for quantifying joint attention in naturalistic interactional contexts.

At the heart of this methodological framework lies the concept of joint attention, a foundational element in early social and cognitive development. For an attentional act to qualify as joint attention, it is not enough for two individuals to focus on the same object simultaneously; they must do so together, as a "plural subject" (Gilbert, 2007). This shared process is primarily mediated through visual cues, consequently for children with VI, where gaze cues are limited, head orientation becomes a critical channel for intersubjectivity and compensatory strategies that support joint attention (Damen et al., 2015). A key aspect of early intervention for VI children involves structured approaches to cultivating joint attention from infancy, including techniques like "shared discovery" to promote mutual exploration and play (Dale & Salt, 2007). This work also reinforces and expands upon long-standing intuitions about the role of eye contact in social interaction. Early interventions aimed at teaching VI children to orient their faces toward speakers, intended to signal attentiveness and conform to visual norms (Foxx, 1977). These early intuitions anticipated the communicative significance of head orientation, which is now increasingly recognized in both clinical and developmental research as a meaningful channel of social engagement beyond vision. This work formalizes long-standing clinical intuitions about the communicative role of head orientation into a quantitative framework, bridging developmental theory and computational modeling.

Due to the arbitrary nature of the angle classifications to discriminate among attentional categories, some considerations are warranted to solidify the present methodological framework. Head orientation was quantified through a vector-based approach, and two computational models (WLSM and GMM) were tested to improve the robustness of attentional category classification beyond arbitrary angle cut-offs. Agreement across clinical coding, computer vision, and motion capture was then evaluated, with GMM showing the highest consistency and being selected for subsequent joint attention analyses. Importantly, the computational models employed in this study were not intended as predictive machine learning tools, but rather as structured ways to reduce arbitrariness in attentional classification and to preserve the continuity of angular information. Both WLSM and GMM were used to model the relational structure of head orientation signals. This makes their application suitable for small numerosity, high-resolution datasets, where interpretability and semantic grounding are prioritized over prediction.

From a computational perspective, conceptually the WLSM aligns with recent trends that incorporate continuous modeling and similarity-based weighting, such as Gaussian Modulated Weighting (GMW), which replaces dot-product similarity with Gaussian kernels to capture graded relationships (Chen et al., 2025). While GMW operates on high-dimensional feature embeddings, the present method applies a similar principle to interaction angles, using Gaussian weighting to model proximity to predefined attentional cutoff. This adaptation enables interpretable, semantically grounded classification while preserving the flexibility of continuous modeling. Moreover, recent approaches to head pose and gaze estimation increasingly favor regression-based models over direct classification, as they better preserve the spatial continuity and semantic proximity between angular configurations. Notably, both the Multi-Loss framework for head pose estimation (Ruiz et al., 2018) and the end-to-end CNN model (Ewaisha et al., 2020) for driver gaze prediction adopt strategies that combine continuous angle prediction with post-hoc discretization or binning, aiming to enhance robustness and interpretability. These methods share conceptual ground with the statistical approach adopted in the present study. In addition to Gaussian weighting, the present work incorporates a GMM to capture the probabilistic structure of head orientation patterns. Unlike deterministic cutoffs, GMM models the distribution of angular configurations as a mixture of Gaussian components, allowing flexible boundaries. This probabilistic approach complements the WLS model by introducing a data-driven layer that accounts for variability without sacrificing interpretability. Unlike clustering techniques that lack semantic anchors, GMM in this context is constrained by relational categories ensuring cognitive plausibility while leveraging statistical rigor.

While the present findings offer valuable insights, several computational and methodological limitations should be acknowledged. The generalizability of the present approach remains constrained by the relatively small dataset. With larger samples, the use of data-driven models such as GMM could extend the interpretability and robustness of the findings, enabling more nuanced insights into variability across interactional contexts and populations. Although the CV pipeline provides a practical and marker-less solution to estimate head orientation, it presents several methodological limitations when compared with the motion-capture system used as gold standard. First, CV estimates yaw, pitch, and roll independently but ultimately reduces these parameters to a 2D head-direction vector in the image plane. This projection entails an inherent loss of depth information and makes the estimate sensitive to out-of-plane rotations, potentially attenuating subtle directional differences. Second, the CV algorithm operates at a lower temporal resolution than the motion-capture system, and its accuracy depends critically on the visibility of facial keypoints. During the experiment, transient occlusions, particularly around the gate structure, occasionally prevented the algorithm from detecting the key landmarks required to build the directional vector. In these instances, the head-direction estimate could not be reliably computed, leading to gaps or reduced precision in the resulting behavioral coding. Nonetheless, the methodological consistency across CV and motion capture proved sufficiently robust, as evidenced by the convergent patterns observed, supporting its validity in capturing meaningful aspects of attentional engagement.

Finally, recent operational definitions of joint attention emphasize the importance of behavioral markers of mutual awareness, such as communicative signals that indicate shared attentional focus between interaction partners (Graham et al., 2021). In this light, an important limitation of the present approach should be acknowledged. While our measurement technique captures the coordination of head orientation in naturalistic settings, it does not, on its own, provide a direct cue to the underlying attentional state or to the presence of mutual awareness. As such, the extent to which head alignment alone can be taken as a valid proxy for the psychological construct of joint attention remains open to interpretation. Experimental paradigms that incorporate more explicit attentional cues or elicitation procedures may offer clearer inferential leverage in this respect.

At the same time, beyond the present experimental setting, the proposed algorithms are readily transferable to other interactional contexts in which head orientation can be estimated, including free play, rehabilitation sessions, and peer-to-peer exchanges. Because the framework relies on continuous head orientation signals rather than task-specific cues, it is adaptable to a wide range of experimental designs and ecological settings, as well as to different developmental populations. In this respect, the proposed metric could be meaningfully integrated within the free play section of the Autism Diagnostic Observation Schedule (ADOS), which is routinely used to assess the severity of autism spectrum symptomatology, providing an objective, fine-grained complement to observational ratings of social attention and engagement. Crucially, in such contexts, head coordination measures should be interpreted as one component within a broader constellation of behavioral indicators, rather than as a standalone measure of joint attention.

While the present study focused primarily on nonverbal cues such as head orientation, an important avenue for future research involves the integration of multiple sources of information that more directly index attentional and communicative states. Verbal exchanges, gaze cues, and other social signals, which often co-occur with attentional behaviors, could be incorporated into multimodal attention metrics to enrich the assessment of relational quality and provide a more comprehensive and theoretically grounded understanding of social engagement, beyond the visual channel.

4.7 Conclusion

Enculturated intersubjectivity: head orientation as an index of joint attention

The present findings indicate that quantifying the presence of joint attention episodes using head orientation is a promising alternative to gaze-based measures, capturing subtle relational dynamics. Overall, these results emphasize the importance of flexible temporal criteria and advanced computational models to represent dynamic relational states. Beyond its technical aspects, the approach addresses a critical clinical need: the availability of quantitative measures that complement qualitative observation. Potential applications include developmental research, clinical assessment, and intervention design. By providing continuous, high-resolution data on head orientation and attentional states, this framework enables clinicians to move beyond qualitative observation and quantify joint attention assessment in naturalistic contexts. Technology guided by clinical expertise can thus offer valuable insights into the interpretation of behavioral patterns, paving the way for next-generation interventions and inclusive interaction models.

From a theoretical standpoint, these methodological and clinical implications directly resonate with the developmental framework outlined in the introduction. If development is enculturated, arising through culturally structured caregiver–child practices, then intersubjectivity necessarily reflects these practices. This includes the sensory modalities that a given culture treats as primary channels for coordination. In typical development, caregivers rely heavily on visual norms (e.g., gaze following) to scaffold shared attention, especially under conditions of uncertainty. When visual cues cannot be relied upon, as in congenital VI, interactional practices reorganize: other body-based signals, particularly head orientation, become primary channels for achieving coordination and constructing shared meaning. In populations with sensory impairments, quantifying these body-based cues is not only clinically valuable but theoretically essential for understanding how children and caregivers negotiate meaning, maintain connection, and enact joint attention within the sensory worlds they inhabit.

Chapter 5

Enabling intersubjectivity as proprioception.

Variability matters, effects of sonification in supporting spatial orientation in the absence of vision

Summary

This chapter explores how enabling technologies mediate intersubjectivity by augmenting proprioceptive feedback. This was done by using the novel framework for Sonification of Navigation and Interactive Activities (SONIA), a sonification system designed to render motor intent audible and support spatial orientation when visual guidance is absent. Specifically, this leverage SONIA to investigate whether auditory cues can improve straight-line walking in children without vision. Ten children with VI and eleven children with TD blindfolded walked along a predefined path under two conditions: with and without real-time auditory feedback. Movement was tracked using a Vicon motion capture system integrated with SONIA, a framework that triggers sound when participants remain within a virtual auditory beam. Trial-level trajectory dispersion was quantified using the Median Absolute Deviation (MAD). A linear mixed-effects model with Group (TD, VI), Condition (OFF, ON), and their interaction as fixed effects, and random intercepts for participants, revealed a significant main effect of Condition, indicating lower MAD in the ON relative to the OFF condition. This effect was further qualified by a significant Group × Condition interaction. Estimated marginal means indicated a clear reduction in MAD for VI children in the ON condition, whereas no systematic change emerged for TD children. Trial-by-trial visualizations corroborated this pattern, showing a consistent separation between OFF and ON trajectories across trials in the VI group and largely overlapping trends in the TD group. Overall, the findings indicate that real-time sonification stabilizes motor performance specifically in children with VI, reducing variability of trajectory execution when visual information is congenitally absent.

5.1 Introduction

5.1.1 Proprioception, agency and sensory feedback in motor control

In 1913, Charles Sherrington introduced the concept of proprioception, describing it as the perception of internal signals arising from the organism's own movements (Sherrington, 1913). Often referred to as the "sixth sense," proprioception provides awareness of body posture and motion, playing a crucial role in everyday tasks (Tuthill & Azim, 2018). Much of human interaction with the world involves producing movement (Han et al., 2016), and in this context, proprioception is fundamental for regulating motor actions.

One influential account of the relationship between proprioception and movement is optimal feedback control, which posits that proprioceptive input supports internal models predicting the sensory consequences of movement, enabling precise planning and execution (Wolpert et al., 2011). Voluntary action unfolds through preparation, motor command, and sensory feedback, integrated by predictive mechanisms that sustain the sense of agency: the capacity to control actions and their effects (Haggard & Eimer, 1999). Phenomena such as intentional binding reveal agency as a cognitive process linking actions and outcomes over time (Haggard et al., 2002). Philosophical perspectives emphasize movement as a continuous, goal-directed stream (Dewey, 1896; Merleu-Ponty, 1962), suggesting that fragmented cues may obscure essential aspects of perceptual experience (Rodger & Craig, 2016). A continuous framework is therefore critical for understanding motor control.

Within this context, sonification, the conversion of movement-related data into non-verbal sound signals, offers a promising approach. By mapping physical and physiological information onto acoustic properties, sonification provides real-time access to biomechanical details that would otherwise remain hidden (Dubus & Bresin, 2013; Effenberg et al., 2011; Pizzera & Hohmann, 2015). By rendering silent phases audible, sonification has been applied in sports to refine motor skills and in rehabilitation to enrich sensory feedback and support recovery (Effenberg et al., 2016; Nown et al., 2022). Auditory cues can function as affordances, invitations to action (Gibson, 1979), and contribute to shaping internal motor representations (Wolpert et al., 2011).

This approach becomes particularly relevant when vision, the primary channel for spatial orientation, is absent. The compromise of visual input affects spatial cognition: the ability to represent relationships between oneself, the environment, and others, leading to delays in orientation and adaptive skills (Bigelow, 1992; Dale & Salt, 2007). These difficulties are closely linked to motor development; for instance, maintaining a straight-line walking trajectory requires the integration of balance control, spatial orientation, and sensorimotor coordination. When visual information is unavailable, the nervous system compensates by reweighting alternative sensory inputs, relying more heavily on proprioceptive, auditory, and vestibular cues (Kallie et al., 2007). Evidence from balance research further shows that auditory cues can support postural control when vision is temporarily withdrawn, modulating self-motion perception and stabilizing motor behavior across different audio conditions (Guarischi et al., 2024). Although these findings derive from studies in sighted individuals under blindfolded conditions, they illustrate how the loss of visual input immediately increases reliance on auditory and proprioceptive information. When vision is not just temporarily but chronically absent, as in children with visual impairment (VI), these compensatory demands become part of everyday motor functioning and naturally extend to locomotion. A recent narrative review documented gait adaptations in children with visual impairment (VI), including slower walking speed, shorter strides, and longer double-support phases (Montagnani et al., 2024). Balance deficits are also well documented, with increased postural sway and altered weight distribution compared to sighted peers (De Araújo et al., 2014; Walicka-Cupryś et al., 2022).

Given these challenges, auditory and proprioceptive cues become essential for spatial understanding and effective movement (Cappagli et al., 2017), and technological interventions can complement traditional mobility training: for example, early audio-motor training using the Audio Bracelet for Blind Interaction (ABBI) has been shown to improve mobility and body awareness in children with VI (Cappagli et al., 2019). Motor behavior, especially during development, is inherently variable, and the structure of this variability provides a sensitive index of sensorimotor control: excessive or unstable dispersion reflects poorer calibration, whereas organized variability supports adaptation and skilled performance (Dhawale et al., 2017). Accordingly, trial-level trajectory variability offers an informative window onto how children regulate movement when vision is unavailable.

In an effort to move the meaning of sounds out of the head and into the environment (Steenson & Rodger, 2015), real-time sonification offers a means of enhancing proprioceptive processing when visual information is absent. By providing continuous auditory feedback contingent on body position, sonification creates a closed sensorimotor loop in which movements generate sound and sound, in turn, informs subsequent movement adjustments. This dynamic coupling may reduce proprioceptive drift, strengthen body–space representations, and improve the accuracy of internal models used for motor control.

Building on these principles, the present study investigated whether real-time sonification enhances proprioception and, consequently, spatial cognition, leading to more accurate motor performance as reflected by reduced trajectory variability. To this end, the straight-walking performance of blindfolded typically developing (TD) children and children with visual impairment (VI) was compared with and without sonification.

5.2 Method

5.2.1 Participants

Data from 11 TD and 10 VI children were analyzed. TD participants (mean age ± SD: 8.82± 1.91 years) were recruited at the Fondazione Istituto Italiano di Tecnologia (Genoa, Italy). Parents or legal guardians provided written informed consent prior to participation. The study protocol was approved by the local Ethics Committee (Ethical Committee, ASL 3, Genoa, Italy).

Children with VI (visual acuity > 1.3 on the logMAR scale; mean age ± SD: 9.2± 2.7 years) were recruited among patients in care at the Fondazione IRCCS Istituto Neurologico Mondino (Pavia, Italy). Inclusion criteria required the absence of neurological, vestibular, cognitive, or psychiatric disorders. The protocol was approved by the local Ethics Committee (Fondazione IRCCS Policlinico San Matteo, Pavia, Italy), and written informed consent was obtained from all parents or legal guardians.

5.2.2 SONIA framework

SONIA is an automated framework that defines spatial paths and sonifies them in real time using motion capture data. The agent's position is continuously tracked, and its relative location with respect to a predefined path governs sound playback: when the agent is within path boundaries, the sound is active; when outside, the sound is muted. This design enables a direct, contingent coupling between movement and auditory feedback.

Real-time position detection relies on the Vicon DataStream SDK for Matlab. The kinematic model is configured in Vicon Nexus (version 2.10), and a single marker is specified by label for real-time tracking. Marker labels are defined by a unique name and color to ensure unambiguous identification. To preserve label integrity, each marker is required to always remain visible to at least three motion-capture cameras. Marker labels and their three-dimensional coordinates are streamed from Nexus to MATLAB via the SDK. If a marker is not detected, positional information is temporarily unavailable; in those frames, the framework cannot evaluate whether the marker lies within the path boundaries, no sound is generated, and the process advances to the next frame.

The position-check and sound-generation loop exhibits a measured average latency of 25 ms ± 7 ms when tested at a motion-capture sampling frequency of 100 Hz. The framework can automatically start and stop motion data acquisition, providing a flexible and responsive setup for studying movement–sound interactions. Automation was implemented in MATLAB R2020b, using a custom USB/serial trigger connected to the main PC and to the Vicon Lock+ control unit (Figure 5.1). If a remote trigger is not available, acquisitions can be initiated manually within Nexus.

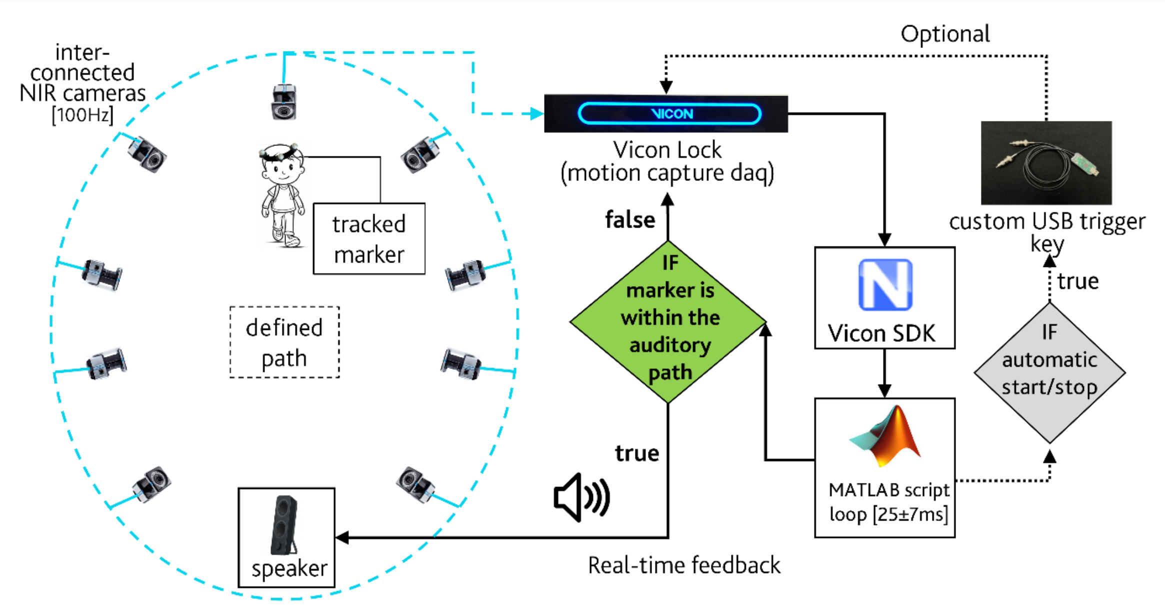

Figure 5.1. Scheme of SONIA framework. On the left, the experimental setup is shown, where the participant walks while being tracked by the Vicon motion capture system. The environment also includes the virtual auditory beam used in the study (represented by the dashed rectangle). Marker positions are detected by the Vicon cameras and transmitted via the Vicon DataStream SDK to MATLAB, which determines whether the marker lies inside or outside the auditory beam and activates the speaker accordingly. The dotted arrows illustrate the optional module for automatically starting and stopping each motion-capture trial. This functionality was implemented through a custom USB trigger key based on a programmable MicroPython board. In this framework, the device operates as a virtual COM port under Windows and produces two independent TTL triggers: acquisition start and stop are controlled via the positive and negative digital edges, respectively, within a 0–3.3 V range.

To facilitate standardized procedures, the framework loads a CSV file containing an example experimental protocol in which auditory conditions (sound ON/OFF) are balanced across a predefined number of participants and trials. At the beginning of each session, the user is prompted to enter the participant ID, and the corresponding sequence of conditions is automatically loaded. An optional training phase is available to allow users to test and familiarize themselves with the auditory path before data acquisition begins. During collection, the framework displays the real-time position of the tracked marker within the calibrated environment for monitoring and quality control. Once the path is created and the motion-capture environment meets the visibility requirements, data acquisition can be started.

5.2.3 Experimental protocol

In the present study a straight auditory beam was implemented to assess walking, stability, and spatial orientation without vision. Participants were blindfolded and asked to walk in a straight line starting from a fixed position. They were informed that they would complete 10 trials in two conditions (ON and OFF), presented in randomized order. In both conditions, their task was to walk as straight as possible along an imaginary path. At the end of each trial, participants were guided back to the starting point by the experimenter.

In the ON condition participants received real-time auditory feedback delivered through a JBL speaker connected to an HP computer. The cue consisted of a brief, moderately loud 480-Hz beep; participants were instructed to walk as straight as possible and to rely on the sound as an anchor, noting that the feedback stopped whenever they drifted off the beam. Each participant completed 10 trials per condition (ON, OFF), starting from a fixed point; condition order was randomized.

The beam width was 20 cm, based on average pediatric step width (McKay et al., 2017), providing a stability margin. The beam was set in a 5.1 × 4 m area. Data were collected at 100 Hz with nine Vicon Vero cameras at Italian Institute of Technology, Genova and eight Vero cameras at the Mondino Foundation, Pavia. To ensure comparability of the spatial metrics across the two recording environments, the calibration was centered on the same coordinate reference in both laboratories, guaranteeing an equivalent spatial alignment and the same effective length of the auditory tunnel. Figure 5.2 illustrates the path used during the experimental session, along with an example of marker detection across different trials.

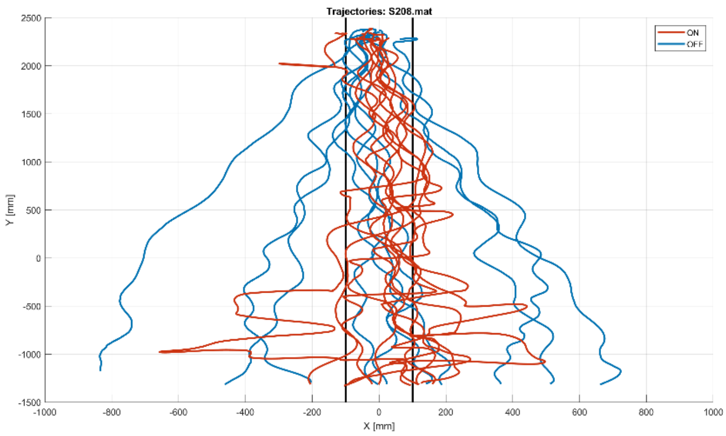

Figure 5.2. Trajectories of an example subject. Auditory beam boundaries (black lines) created in the physical environment. Real-time marker trajectories are shown for trials in the ON condition (orange) and OFF condition (blue).

5.2.4 Data processing

During acquisition, continuous marker visibility was ensured along the beam. Trajectories were filtered in Vicon Nexus with a low-pass Butterworth filter (cut-off 6 Hz), and .c3d files with x, y, z of the central marker were exported for analysis.

5.2.5 Statistical analysis

Statistical analyses were carried out using R (version 4.2.2). The dependent variable was the Median Absolute Deviation (MAD) of each trajectory, representing trial-level movement dispersion. To account for repeated measurements and the clustering of multiple trials within each participant, a linear mixed-effects model was fitted. The model included Group (TD vs. VI), Condition (ON vs. OFF), and their interaction as fixed effects, and a random intercept for participants to capture individual variability. To clarify interaction effects, estimated marginal means were computed, followed by pairwise comparisons between conditions within each group and between groups within each condition.

5.3 Results

5.3.1 Findings

The linear mixed-effects model revealed a significant main effect of Condition, with lower trajectory dispersion in the ON condition compared to the OFF condition ($b = -32.71$, $t(397) = -3.40$, $p < .001$). The main effect of Group was not significant ($b = -30.22$, $t(23.06) = -1.38$, $p = .18$). A significant Group × Condition interaction was observed ($b = 27.81$, $t(397.03) = 2.09$, $p = .037$), indicating that the effect of Condition differed across groups.

Follow-up comparisons showed that children with VI exhibited a significant reduction in MAD from OFF to ON ($t(397) = 3.40$, $p < .001$), whereas no condition-related change was found in the TD group ($t(397) = 0.53$, $p = .59$). Between-group comparisons were non-significant in both the OFF condition ($t(23.1) = 1.38$, $p = .18$) and the ON condition ($t(23.0) = 0.11$, $p = .91$) (Figure 5.3).

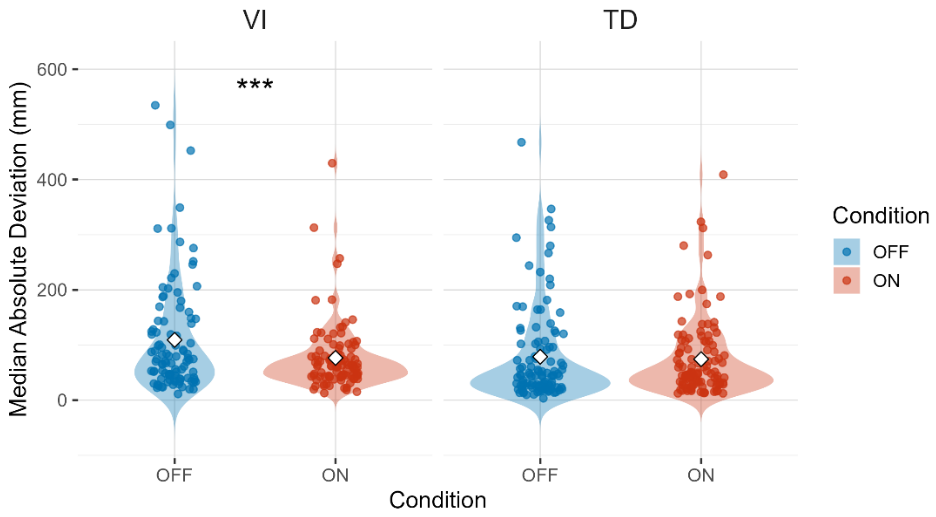

Figure 5.3: Distribution of trial-level Median Absolute Deviation for each group (VI, TD) across the two experimental conditions (OFF, ON). Violins illustrate the density of observations, individual points represent trial values, and the white diamond denotes the mean for each condition.

To examine whether performance evolved systematically within the experimental session, trial-by-trial trajectories of MAD were considered. A qualitative inspection of these trajectories indicates that the separation between the OFF and ON conditions is consistently maintained across trials in the VI group, whereas in the TD group the two conditions display largely overlapping patterns with minimal divergence over time (Figure 5.4).

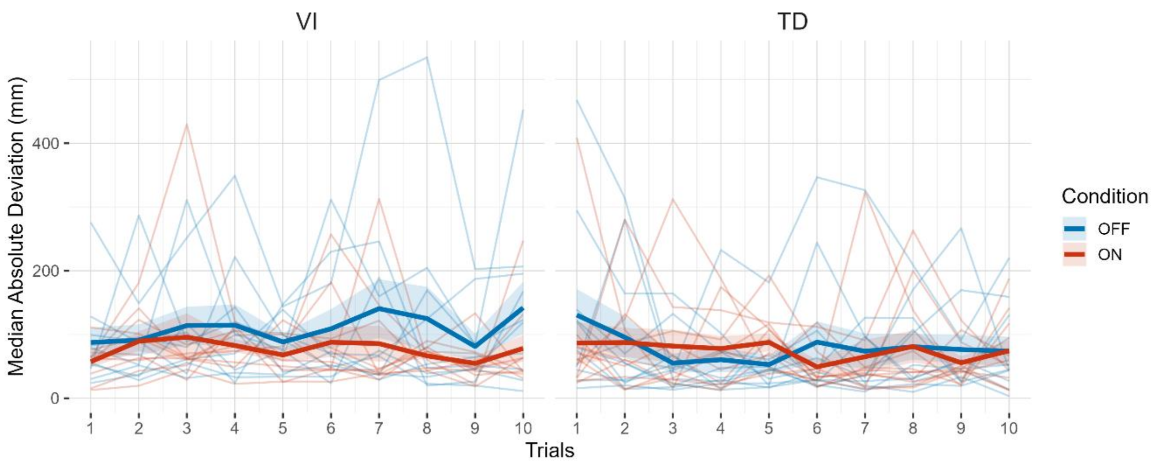

Figure 5.4: Trial by trial trajectories of Median Absolute Deviation (MAD) for each group (VI, TD) across the two experimental conditions (OFF, ON). Lines represent group-level mean MAD values across aligned trials, and shaded ribbons indicate ±SE.

5.4 Discussion

The present work support the hypothesis that real-time sonification can enhance proprioceptive processing and spatial control when vision is absent (Hermann et al., 2005). By providing an additional sensory channel, auditory feedback likely complemented proprioceptive input, enriching internal models for movement prediction and error correction and reducing reliance on proprioception alone (Scott, 2004). This suggests that sonification works synergistically to strengthen sensorimotor integration when visual cues are unavailable.

The findings indicate an influence of auditory feedback on trial-level trajectory dispersion, specifically reflected in a reduction of MAD. The mixed-effects model revealed a significant main effect of Condition, with overall lower MAD values in the ON condition compared to the OFF condition. However, this effect was qualified by a significant Group × Condition interaction, indicating that the influence of sonification differed between VI and TD participants. Follow-up contrasts showed that children with VI exhibited a clear reduction in MAD when auditory feedback was provided, whereas TD children showed no systematic change across conditions. This pattern was further supported by the trial-wise visualizations, in which the separation between the OFF and ON trajectories remained consistently visible across trials in the VI group, but largely overlapped in the TD group.
These findings suggest that auditory feedback contributed to stabilizing motor output specifically in children with VI, reducing trial-to-trial fluctuations in trajectory execution. In contrast, TD participants, who likely relied on more established proprioceptive and visuomotor priors, did not benefit from the additional auditory cues in terms of dispersion reduction. Taken together, the results indicate that real-time sonification does not uniformly affect all participants but exerts its strongest impact when visual information is congenitally absent, enhancing consistency and reducing variability in movement execution for children with VI.

These results can be situated within Bernstein's principle of repetition without repetition, according to which motor learning does not consist in reproducing identical movements but in organizing the redundancy of degrees of freedom more effectively; variability is not merely noise but a resource for exploration and adaptation (Bernsteĭn, 1967). Motor variability it has been considered a noise that has to be deleted, alternative perspective suggests that it may be a feature of how sensorimotor circuits operate and learn (Herzfeld & Shadmehr, 2014). Indeed, alternative accounts echo this view, emphasizing that controlled variability supports behavioral flexibility and skill acquisition (Newell & Corcos, 1993; Ranganathan et al., 2020). In this light, the reduction of dispersion observed in VI under auditory feedback may reflect a shift from uncontrolled to functional variability, consistent with the idea that enriched non-visual feedback can help stabilize coordination patterns when vision is congenitally absent.

This study presents some limitations. First, the relatively small sample size restricts the generalizability of the findings. Future research should involve larger cohorts and investigate developmental trajectories, examining whether age influences motor performance and responsiveness to auditory feedback. Second, this work represents an initial implementation of the SONIA system. Future developments of the SONIA system could move beyond simple ON/OFF cues toward continuous and interactive sonification, allowing a more fine-grained coupling between bodily movement and auditory feedback. Rather than providing discrete signals, sound could be mapped in an error-sensitive and graded manner to body orientation and trajectory deviation (e.g., through changes in pitch, timbre or spatialization), thereby supporting anticipatory adjustments and online motor correction. Such an extension would enable the application of sonification-based feedback to more complex locomotor tasks, including navigation in environments with obstacles or dynamically changing spatial constraints. Importantly, this approach should not be understood as an alternative to established non-visual navigation strategies such as echolocation or sensory substitution devices, which primarily convey information about external spatial layout and object presence. Instead, the present framework targets internal body–space calibration, reinforcing proprioceptive control and motor consistency by shaping the child's internal model of self-motion. In this sense, auditory feedback functions as an embodied regulatory signal rather than an explicit spatial representation. This distinction is particularly relevant in rehabilitation contexts, where fostering stable and adaptive sensorimotor organization is a prerequisite for effective interaction with more complex perceptual tools. By promoting a dynamic and continuous exchange between the child and the technological system, such an approach may emulate the contingent interactions typically observed between clinicians or caregivers and children, thereby enhancing ecological validity and situating sonification as an intersubjective scaffold rather than a purely assistive device.

5.4 Conclusion

Enabling intersubjectivity: real-time sonification stabilizes motor performance

Real-time sonification enhanced trajectory accuracy and reduced variability, particularly in VI children. These findings suggest that auditory feedback can serve as an effective compensatory cue for spatial orientation and motor control when vision is compromised, supporting the integration of sonification into mobility training and rehabilitation programs. More broadly, treating motor variability not as mere noise but as a meaningful signature of sensorimotor organization underscores its value as a sensitive marker of motor function, especially in clinical populations, where changes in variability can reveal subtle differences in control, adaptation, and developmental trajectories.

These findings gain additional significance when viewed through the developmental framework introduced in the introduction. Development is enabling: it expands the child's perceptual and relational possibilities, opening new avenues for exploration and coordination. Sonification can be understood within this logic. By turning movement into sound, it introduces technological affordances that reorganize the available information and offer contingent cues that enhance bodily awareness and support spatial navigation. For children with VI, such auditory information can function as a scaffold within the zone of proximal development, providing real-time guidance that refines precision and coordination when vision cannot serve as the primary source of environmental access. If development extends the space of perceptual and relational possibilities, then intersubjectivity can also be enabled through assistive technologies. Sonification offers a channel through which coordination and co-reference can be achieved in forms that remain contingent and accessible to the sensory conditions of children with VI.

Chapter 6

General discussion

This thesis examined how visual experience shapes intersubjective dynamics in early development by integrating four complementary dimensions: embodied, embedded, enculturated, and enabling. Across studies, the findings suggest that when visual input is reduced, interaction and movement patterns might reorganize in context-dependent ways. Autonomic responses showed variation with caregiver identity, motor variability appeared to act as a potential resource under uncertainty, head orientation emerged as a potential proxy measure of joint attention beyond gaze, and auditory cues showed promise in supporting spatial orientation. Together, these results invite a reconsideration of strictly vision-centric accounts of early social engagement and introduce a set of emerging metrics and tools to detect and scaffold interaction in VI populations.

Embodied intersubjectivity: disruption of caregiver-specific autonomic tuning under visual deprivation

The first study examined whether caregiver identity modulates autonomic responses during controlled tactile stimulation under visual deprivation, adapting a previous protocol (Aguirre et al., 2019). Although no significant main effects of Group or Caregiver Identity emerged, a significant interaction suggested diverging tendencies between TD and VI infants. TD infants showed a pattern broadly consistent with previous literature: lower heart-rate indices when tactile stimulation plausibly originated from the mother compared to a stranger. Instead, infants with VI did not exhibit this differentiation. These trends may hint at a reorganization of caregiver-specific physiological tuning when visual input is congenitally reduced or unavailable. Building on this, visual deprivation could attenuate the integration of identity cues during social touch, leading infants with VI to rely on tactile information in ways that differ from TD infants. Such reliance might enhance sensitivity to certain tactile features while reducing the ability to map them onto caregiver identity, thereby altering the autonomic signature typically associated with maternal touch.

Importantly, these patterns resonate with prior findings showing that social touch is a strong regulator of infant physiology. Maternal tactile contact is known to decrease arousal, stabilize autonomic state, and buffer stress responses in TD infants (Feldman et al., 2002; Feldman et al., 2011). Similar benefits have been reported in preterm populations, where tactile and kinesthetic stimulation improves physiological regulation, pain responses, and homeostasis (Manzotti et al., 2019). In this broader context, the present findings that VI infants did not show the expected maternal-related cardiac attenuation could reflect a disruption in the recognition of the touch source. Taken together, these observations provide a contribution to accounts of embodied intersubjectivity: while tactile input remains salient for physiological regulation, its integration with identity cues seems to depend on multimodal contingencies that are altered when vision is absent.

Embedded intersubjectivity: variability and proximity to caregivers under multisensory uncertainty

The second study explores the embedded dimension into an ecological setting inspired by the Strange Situation paradigm (Ainsworth et al., 1978), examining how unexpected multisensory cues modulate kinematics and proximity to caregivers in children with VI. Interestingly, children with VI showed larger variability within their motor trajectory and kept a shorter distance from the caregivers when multisensory cues were introduced in the playroom. Interpreted within an attachment framework (Bowlby, 1969), these patterns suggest that under novelty and uncertainty, children with VI seems to recruit proximity seeking as a regulatory strategy and flexibly reorganize motor exploration to regain control over the environment. This is consistent with accounts that link attentional flexibility and problem-solving to the balance between attachment and exploration (Diamond & Fagundes, 2008; Gross, 2013; Main, 2000; Pallini et al., 2019) and with evidence that vision supports early attention control (Cass et al., 1994; Dale & Sonksen, 2002; Rothbart et al., 2006; Tadić et al., 2009). In line with ecological approaches to development (Adolph et al., 2019; Dahl, 2017; Tamis-LeMonda et al., 2017) the study underscores that motor behavior is relational and context-sensitive. Crucially, the quality and predictability of stimulation appear decisive: previous work shows that auditory cues can enhance shared attention to objects while potentially interfering with peer social interaction depending on context (Verver et al., 2019). Converging with clinical observation, the present data suggest that as visual acuity decreases, new sensory inputs might initially inhibit exploration; hence predictable, graded multisensory supports might be essential for promoting adaptive regulation (Caron et al., 2023; Gui et al., 2023; van den Broek et al., 2017a).

Enculturated intersubjectivity: head orientation as an index of joint attention

The third study formalized long-standing clinical intuitions into a quantitative framework: head orientation could index joint attention during naturalistic mother–child interactions (Beebe & Stern, 1977; Brazelton et al., 1974). Building on these insights, the present study introduced a method for calculating attentional vectors derived from head orientation data in caregiver–child dyads. The general aim was to develop method for detecting joint attention episodes during ecological mother–child interactions using clinical video coding, CV and marker-based motion tracking. Agreement in attentional category classification was high between computer vision (CV) and motion capture and highest when angles were classified with a Gaussian Mixture Model (GMM), while lower concordance with clinical coding likely reflected temporal misalignment rather than categorical disagreement. Methodologically, the framework advances beyond arbitrary angle cutoffs by integrating two computational probabilistic models (WLSM and GMM). This study contributes a validated, scalable, and non-invasive method for analyzing attentional dynamics in caregiver–child interactions, preserving ecological validity and ensuring reliable data collection. Beyond its technical aspects, the approach addresses a critical clinical need: the availability of quantitative measures that complement qualitative observation. The approach centers the communicative role of head orientation and reframe it as a culturally meaningful, enculturated channel of intersubjectivity.

Enabling intersubjectivity: real-time sonification stabilizes motor performance

The fourth study tested whether real-time sonification could serve as an enabling, compensatory channel for spatial orientation and proprioceptive control when vision is absent. Previous early audio-motor training has been shown to improve mobility and body awareness in children with VI (Cappagli et al., 2019). In the present work auditory feedback seems to enables VI children to organize their movement variability more effectively, a hallmark of adaptive motor control. The sound led to a marked reduction in trajectory dispersion for VI children, with lower MAD values in the ON condition and no corresponding change in the TD group, suggesting that sonification supports the transition to a more stable straight-line walking pattern specifically in cases of congenital visual absence. Framed through Bernstein's repetition without repetition, the results indicate a shift from uncontrolled to functional variability, optimizing degrees of freedom (Bernsteĭn, 1967; Newell & Corcos, 1993; Ranganathan et al., 2020). This enabling effect is clinically promising: well-designed sonification can be ecologically integrated into rehabilitation. Importantly, next-generation systems may benefit from moving to continuous, contingent, and semantically informative mappings that emulate the reciprocity typical of caregiver–child exchanges.

By weaving these results together, the thesis advances a preliminary model of embodied intersubjectivity in which social understanding emerges from multisensory coupling and action. The four E's clarify complementary levels of organization: (i) Embodied, physiological attunement seems to be modulated by altered sensory weighting; (ii) Embedded, motor strategies and proximity regulation seems to vary with contextual uncertainty; (iii) Enculturated, head orientation function as one component of communicative coordination; (iv) Enabling, sonification technology seems to augment perception–action loops, stabilizing motor performance. Rather than treating VI as a deficit against a visual norm, the synthesis highlights possible adaptive pathways including sensory reweighting, functional variability, and alternative communicative channels that require further empirical validation through larger samples, controlled comparisons, and multimodal measurement approaches. More broadly, advancing the field requires revisiting ecological and dynamic theories of development through new methodological lenses, moving from static metrics toward temporally sensitive, contingency-driven approaches capable of capturing intersubjective processes and co-regulation in real time.

6.1 Clinical and technological implications

A key clinical insight emerging from this thesis concerns the finding that infants with VI might show reduced differentiation in responses to the source of social touch. This pattern of caregiver-specific autonomic tuning is consistent with a broader body of research showing that early intersubjectivity relies on the integration of multimodal contingencies that scaffold the infant's ability to differentiate familiar from unfamiliar partners (Feldman, 2012). In typical development, parental contact, voice, and gaze converge to create coherent relational expectations (Ferber et al., 2008), whereas in infants with VI, the absence of vision might limit the degree to which multimodal representations become fully integrated, and feedback from unimodal channels may be less informative. The present findings therefore suggest that touch does not simply compensate for reduced visual access but likely depends on its embedded within predictable multisensory patterns to effectively sustain early social meaning. Clinically, the findings related to motor strategies and proximity regulation in contextual uncertainty highlight the importance of predictable and meaningful multisensory supports, caregiver forewarning, and structured shared discovery to foster joint attention and regulate arousal (Dale & Salt, 2007). Research consistently shows that the temporal structure of caregiver–infant contingencies is a core driver of affective coordination (Beebe & Stern, 1977) and that disruptions in the coherence or timing of such contingencies affect the development of social expectations and self–other regulation (Tronick & Cohn, 1989). In infants with VI, where access to distal cues is limited, these temporal contingencies become even more critical (Bigelow, 2003). Within this framework, objective metrics such as head orientation could complement expert clinical judgment by increasing the sensitivity and reproducibility of assessments. The potential translational value of these metrics extends to other neurodevelopmental conditions in which joint attention is challenged, including ASD (Pring, 2005), reinforcing that early social coordination depends on accessible, continuous cues that support alignment.

Technologically, the work related to joint attention provides a novel, promising and scalable pipeline that integrates clinical coding, CV, and motion capture, while the sonification framework demonstrates how enabling systems can scaffold navigation and bodily awareness. Together, these technologies outline a preliminary coherent framework for assessing and supporting early interaction that is both rigorous and ecologically grounded, bridging the gap between laboratory precision and clinical feasibility. Design principles emerge: the quality and temporal structure of stimuli matter; contingency promotes engagement. A further clinical implication concerns the value of analyzing infant behavior as time series, rather than collapsing it into single mean values. This approach captures the unfolding of the child’s internal dynamics and opens new avenues for examining dyadic processes, particularly caregiver–infant co-regulation over time. Clinically, this represents a shift from static, unitary metrics (such as subjective ratings or overall averages) to temporally sensitive measures that reflect how behavioral states evolve moment by moment. Importantly, clinical observation is never a snapshot of the present alone: it is inherently informed by the child’s recent history while remaining oriented toward anticipated future states. These tools can be integrated in multiple ways when working with children with developmental disabilities in general, and with VI in particular. First, technology can support the investigation in developmental contexts, facilitating research settings that more closely approximate the cross-cutting ecological principles (Bronfenbrenner, 1977). Second, technology can contribute to assessment and intervention, provided that its development remains grounded in empirical validation. This implies adopting a circular model in which empirical findings guide technological design, and technological innovation in turn refines clinical understanding, generating an iterative improvement of both tools and practices. This implies not only adapting existing tools to the specific sensory profiles of children with VI but also rethinking technological design toward systems that become increasingly human-centered, sensitive to contingencies, and responsive to real-time interactional demands. Ultimately, the convergence of clinical insight and enabling technology points toward interventions that do not merely compensate for sensory loss but actively support the emergence of intersubjective engagement through tailored, dynamic, and ecologically valid multisensory experiences.

6.2 Limitations and future directions

Across studies, sample sizes were modest, limiting the generalizability of findings and warranting replication with larger cohorts. For the embodied dimension, this limitation is particularly critical: expanding the sample is essential to confirm whether the absence of caregiver-specific autonomic tuning under visual deprivation is robust. Additionally, the absence of blindfolded conditions for sighted participants limits the extent to which conclusions can be drawn about the respective roles of experiential vision and current visual input. Future studies specifically aimed at disentangling the effects of temporary visual deprivation on social touch will be needed to address this issue.

Regarding the embedded dimension, limitations include the absence of a parallel video-coding counterpart and the continuous alternation between multisensory and silent conditions, which may have constrained the consolidation of reunion behaviors. A block design might have allowed these effects to stabilize over time; however, the primary aim of the present study was to capture average behavior within trials rather than temporal dynamics across extended exposure. Future work should combine automated kinematic quantification with manual coding to validate relational interpretations, restructure conditions to allow longer familiarization with sensory stimuli, and explicitly consider the temporal dimension across the full duration of the experiment.

In the enculturated dimension, while a multidimensional approach was adopted, joint attention coding would benefit from computational advancements and multimodal integration, incorporating linguistic markers and affective vocalizations alongside head orientation. Expanding the dataset will enable more robust probabilistic modeling and improve temporal alignment between clinical and automated annotations. However, it remains debatable whether the measurement technique employed in a naturalistic setting can fully capture the psychological construct of joint attention: while head coordination provides an index of orienting behavior, the absence of converging cues to attention makes its interpretation less straightforward. Within this perspective, structured paradigms could offer a more direct operationalization of joint attention, providing clearer links between observable behavior and the underlying construct. Potential applications of this framework extend beyond VI to developmental research and clinical assessment in conditions such as ASD, where joint attention is similarly challenged (Pring, 2005). More structured adaptations of this approach could be particularly relevant for standardized assessment contexts, such as the free play section of the ADOS, where joint attention is explicitly evaluated, offering a more direct and clinically interpretable operationalization of the construct.

For the enabling dimension, the current sonification system represents an initial implementation. Advancing toward continuous and contingent mappings with semantically meaningful cues will better emulate the reciprocity of caregiver–child interaction and enhance ecological validity. Future designs should explore adaptive protocols that respond dynamically to movement, fostering interactive exchanges rather than static feedback.

Taken together, these limitations highlight the need for ecologically valid studies that integrate multimodal data and adaptive technologies**.**

6.3 Conclusion

This thesis advances the state of the art by addressing a long-standing limitation in the literature: the predominance of vision-centric models, which have traditionally treated gaze as the primary, often the exclusive, indicator of engagement and joint attention. Such an approach has made it difficult to fully understand early social interaction in children with VI, both because alternative channels were theoretically marginally addressed and because existing tools were not designed to measure them. By contrast, this work suggests that when vision is compromised, intersubjectivity might be supported through alternative pathways that the current literature had only partially acknowledged but never systematically quantified.

Through the integration of physiological measures, kinematic analyses, head-orientation modelling, and movement sonification, the thesis introduces novel and promising ways to capture dimensions of interaction in more ecologically grounded context. The findings suggest that touch might operate as a selective regulatory system, proximity and motor variability could function as adaptive strategies for managing uncertainty, head orientation offers a reliable non-visual marker of shared attention, and auditory feedback could compensate for the absence of visual cues by stabilizing orientation and supporting co-reference. In doing so, the thesis extends current frameworks by shifting the interpretation of early interaction from a deficit-focused account, rooted in what vision cannot provide, to an adaptive, dynamic, and multisensory understanding of what children with VI do provide through reorganized bodily and auditory resources.

Conceptually, the thesis enriches the field by shifting the operational focus of the 4E framework from a general account of developmental processes to intersubjectivity itself, using it as a conceptual lens through which the embodied, embedded, enculturated and enabling dimensions can be observed without mediation by the visual channel. By examining intersubjectivity in conditions where vision is absent, the work highlights how these four dimensions can manifest through alternative sensory modalities and how the interpersonal bond emerges as a multisensory exchange grounded in amodal qualities such as rhythm, contingency and temporal alignment. In this way, the thesis positions intersubjectivity as a privileged window onto the multisensory nature of early social experience, making visible forms of coupling that remain in the background when vision dominates the communicative landscape. In linking these elements back to the initial gap, a research tradition constrained by visual assumptions, the thesis offers a coherent alternative grounded in adaptation, multisensory coupling, and measurable behavioral signatures, suggesting potential avenues for inclusive assessment and rehabilitation intervention that allow the presence of the “other” to emerge and be co-constructed through channels that extend outside the visual domain.

6.4 Postface: the hidden dimension of connection, a phenomenological view

A phenomenological perspective offers a broader interpretive horizon for understanding early intersubjectivity, situating multisensory engagement within an account of experience in which relationality precedes the subject. In Husserl’s writings, the encounter with the other arises through *Paarung*, a pre-reflective pairing in which one’s own embodied subjectivity resonates with the lived body of another (Husserl, 1952). Although this formulation still presupposes a constituting “I,” it already places alterity at the level of immediate sensorial givenness, prior to conceptual or representational differentiation.

Heidegger advances this insight by questioning the notion of a self-contained subject who subsequently enters into relation. Human existence is, from the outset, *Mit-Dasein*: a *being-with-others* (Heidegger, 1927). Intersubjectivity is therefore not a psychological milestone but an ontological structure: the self is not an enclosed interiority but a threshold through which world and others co-belong. Relation does not occur within experience; it is what makes experience possible in the first place.

Merleau-Ponty brings this shift to its fullest expression by grounding relationality directly in the perception. His notion of reversibility, the simultaneity of being sensing and sensible, seeing and being seen, reveals perception as fundamentally intercorporeal (Merleau-Ponty et al., 2013). The senses are not independent channels but variations of a shared expressive field; the other is not inferred but encountered as a point of view within a perceptual dimension that precedes individual subjectivity, returning the self to itself through a dynamic interplay of exposure, resonance, and mutual presence.

Within this phenomenological horizon, intersubjectivity appears not as the product of a particular sensory modality but as a multisensory, rhythmic, and embodied field of co-presence. Vision offers one possible form of reciprocity, yet the relational structure of experience does not depend on it. What is primary is the pre-reflective matrix through which bodies are already entwined; a matrix in which self and other emerge together before they can be distinguished as separate poles.

In this pre-objective depth, the visible world is always underwritten by an invisible layer of intertwining that conditions all perception (Merleau-Ponty, 1968). In this sense, the phenomenological tradition gives conceptual form to what this thesis names the hidden dimension of connection: the relational ground that precedes explicit interaction, exceeds any single modality, and sustains the emergence of self and other within a shared world.

Supplementary materials

Supplementary Table 2.1: Clinical details of the VI participants.

Subject ID	Age (months)	Clinical Diagnosis
S01	11	Nystagmus
S02	8	Septo-optic dysplasia
S03	13	Incomplete isolated aniridia, nystagmus, bilateral mydriasis
S04	8	Leber's congenital amaurosis
S05	11	Leber's congenital amaurosis
S06	8	Nystagmus
S07	10	Nystagmus
S08	10	Septo-optic dysplasia
S09	12	Bilateral corneal leucoma
S10	8	Nystagmus

Supplementary Table 3.1. Demographics information of participants.

	Mean	SD	Min	Max
Age at testing (years)	3.3	1.1	2.0	5.6
Age at confident walking (years)	1.5	0.4	1.0	2.33
Weight (kg)	15.0	0.9	13.0	16.0
Height (cm)	94.8	8.6	75.0	103.0

Supplementary Table 3.2. Clinical information of VI participants.

Subject ID	Age (years)	Clinical Diagnosis	BCVA (LogMar)
S01	3.8	Retinopathy of Prematurity	0.5
S02	3.6	Oculocutaneous Albinism	0.6
S03	2	Oculocutaneous Albinism	0.7

S04	2.8	Oculocutaneous Albinism	0.8
S05	5.6	Leber Congenital Amaurosis	> 1.3
S06	4.5	Inherited retinal distrophy	1
S07	3.11	Inherited retinal distrophy	1
S08	3.3	Bilateral chorioretinal coloboma	1
S09	3.8	Inherited retinal Dystrophy	1
S10	4	Inherited retinal Dystrophy	1.1
S11	4.6	Optic Nerve Hypoplasia	0.5

Supplementary Table 4.3. Clinical information of VI participants.

Subject ID	Age (years)	Clinical Diagnosis	BCVA (LogMar)
S01	3	Albinism	0.7
S02	4	Albinism	0.15
S03	4	Congenital cataract	0.3
S04	5	Albinism	1
S05	4	Congenital cataract	0.7

Supplementary Table 5.1. Clinical details of the VI participants

Subject ID	Age (years)	Clinical Diagnosis	BCVA (LogMar)
S01	4	Congenital retinal dystrophy	> 1.3
S02	10	Congenital retinal dystrophy	> 1.3
S03	10	Congenital retinal dystrophy	> 1.3

S04	10	Congenital retinal dystrophy	> 1.3
S05	11	Congenital retinal dystrophy	> 1.3
S06	12	Congenital retinal dystrophy	> 1.3
S07	13	Congenital retinal dystrophy	> 1.3
S08	9	Congenital retinal dystrophy	> 1.3
S09	6	Ocular structural abnormalities	> 1.3
S10	7	Congenital retinal dystrophy	> 1.3

References

Abney, D. H., Smith, L. B., & Yu, C. (2017). It's time: Quantifying the relevant timescales for joint attention. Proceedings of the Annual Meeting of the Cognitive Science Society,

Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues. *Journal of Child Language*, *44*(5), 1088-1116. https://doi.org/10.1017/S0305000916000416

Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The development of symbol-infused joint engagement. *Child development*, *75*(4), 1171-1187. https://doi.org/10.1111/j.1467-8624.2004.00732.x

Adolph, K. E. (2008). Motor and physical development: Locomotion. In *Encyclopedia of infant and early childhood development* (pp. 359-373). Elsevier Inc. https://doi.org/10.1016/B978-0-12-809324-5.05848-X

Adolph, K. E. (2020). An ecological approach to learning in (not and) development. *Human Development*, *63*(3-4), 180-201. https://doi.org/10.1159/000503823

Adolph, K. E., & Franchak, J. M. (2017). The development of motor behavior. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1-2). https://doi.org/10.1002/wcs.1430

Adolph, K. E., & Hoch, J. E. (2019). Motor development: Embodied, embedded, enculturated, and enabling. *Annual review of psychology*, *70*(1), 141-164. https://doi.org/10.1146/annurev-psych-010418-102836

Adolph, K. E., Hoch, J. E., & Ossmy, O. (2019). James Gibson's ecological approach to locomotion and manipulation: Development and changing affordances. In *Perception as Information Detection* (pp. 222-236). Routledge. https://doi.org/10.4324/9780429316128-14

Adolphs, R. (1999). Social cognition and the human brain. *Trends in cognitive sciences*, *3*(12), 469-479. https://doi.org/10.1016/S1364-6613(99)01399-6

Aguirre, M., Couderc, A., Epinat-Duclos, J., & Mascaro, O. (2019). Infants discriminate the source of social touch at stroking speeds eliciting maximal firing rates in CT-fibers. *Developmental Cognitive Neuroscience*, *36*, 100639. https://doi.org/10.1016/j.dcn.2019.100639

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). Strange situation procedure. *Clinical Child Psychology and Psychiatry*. https://doi.org/10.1037/t28248-000

Alfaro, A. U., Morash, V. S., Lei, D., & Orel-Bixler, D. (2018). Joint engagement in infants and its relationship to their visual impairment measurements. *Infant Behavior and Development*, *50*, 311-323. https://doi.org/10.1016/j.infbeh.2017.05.010

Alfaro, A. U., Vacaru, S., Wittich, W., & Sterkenburg, P. S. (2021). Identifying when children with visual impairment share attention: A novel protocol and the impact of visual acuity. *Infant Behavior and Development*, *64*, 101585. https://doi.org/10.1016/j.infbeh.2021.101585

Als, H., & B. McAnulty, G. (2011). The newborn individualized developmental care and assessment program (NIDCAP) with kangaroo mother care (KMC): comprehensive care for preterm infants. *Current women's health reviews*, *7*(3), 288-301. https://doi.org/10.2174/157340411796355216

Ammaniti, M., & Trentini, C. (2024). The Paths of Intersubjectivity During Infancy. In *WAIMH Handbook of Infant and Early Childhood Mental Health: Biopsychosocial Factors* (Vol. 1, pp. 227-247). Springer.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology*, *36*(2), 190. https://doi.org/10.1037/0012-1649.36.2.190

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press.
Beebe, B., & Stern, D. N. (1977). Engagement-disengagement and early object experiences. In *Communicative structures and psychic structures: A psychoanalytic interpretation of communication* (pp. 35-55). Springer.
Bernsteĭn, N. (1967). The co-ordination and regulation of movements. *PergamonPress Ltd*.
Bigelow, A. E. (1992). Locomotion and search behavior in blind infants. *Infant Behavior and Development*, *15*(2), 179-189. https://doi.org/10.1016/0163-6383(92)80022-M
Bigelow, A. E. (2003). The development of joint attention in blind infants. *Development and psychopathology*, *15*(2), 259-275. https://doi.org/10.1017/S0954579403000142
Bigelow, A. E., MacLean, K., & Proctor, J. (2004). The role of joint attention in the development of infants' play with objects. In: Wiley Online Library.
Bonald, T. (2019). Expectation-maximization for the gaussian mixture model. *Telecom ParisTech. Retrieved*, *12*(14), 2022.
Bornstein, M. H., & Manian, N. (2013). Maternal responsiveness and sensitivity reconsidered: Some is more. *Development and psychopathology*, *25*(4pt1), 957-971. https://doi.org/10.1017/S0954579413000308
Boundy, E. O., Dastjerdi, R., Spiegelman, D., Fawzi, W. W., Missmer, S. A., Lieberman, E., Kajeepeta, S., Wall, S., & Chan, G. J. (2016). Kangaroo mother care and neonatal outcomes: a meta-analysis. *Pediatrics*, *137*(1), e20152238. https://doi.org/10.1542/peds.2015-2238
Bowlby, J. (1969). *Attachment and loss*. Random House.
Braten, S. (2004). Hominin infant decentration hypothesis: mirror neurons system adapted to subserve mother-centered participation. *Behavioral and Brain Sciences*, *27*(4), 508-509. https://doi.org/10.1017/S0140525X0427011X
Bråten, S. (2007). *On being moved: From mirror neurons to empathy*. John Benjamins Publishing. https://doi.org/10.1075/aicr.68
Brauer, J., Xiao, Y., Poulain, T., Friederici, A. D., & Schirmer, A. (2016). Frequency of maternal touch predicts resting activity and connectivity of the developing social brain. *Cerebral Cortex*, *26*(8), 3544-3552. https://doi.org/10.1093/cercor/bhw137
Brazelton, T. B., Koslowski, B., & Main, M. (1974). The origins of reciprocity: The early mother-infant interaction.
Bremner, A. J., & Spence, C. (2017). The development of tactile perception. *Advances in child development and behavior*, *52*, 227-268. https://doi.org/10.1016/bs.acdb.2016.12.002
Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American psychologist*, *32*(7), 513. https://doi.org/10.1037/0003-066X.32.7.513
Brunes, A., B. Hansen, M., & Heir, T. (2019). Loneliness among adults with visual impairment: prevalence, associated factors, and relationship to life satisfaction. *Health and quality of life outcomes*, *17*, 1-7. https://doi.org/10.1186/s12955-019-1096-y
Budman, I., Meiri, G., Ilan, M., Faroy, M., Langer, A., Reboh, D., Michaelovski, A., Flusser, H., Menashe, I., & Donchin, O. (2019). Quantifying the social symptoms of autism using motion capture. *Scientific Reports*, *9*(1), 7712. https://doi.org/10.1038/s41598-019-44180-9
Burkart, J. M., & Southgate, V. (2025). An evolutionary perspective on altercentrism. *Neuroscience & Biobehavioral Reviews*, *176*, 106280. https://doi.org/10.1016/j.neubiorev.2025.106280
Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy*, *1*(2), 149-219. https://doi.org/10.1207/S15327078IN0102_1

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE transactions on pattern analysis and machine intelligence*, *43*(1), 172-186. https://doi.org/10.1109/TPAMI.2019.2929257

Cappagli, G., Cocchi, E., & Gori, M. (2017). Auditory and proprioceptive spatial impairments in blind children and adults. *Developmental science*, *20*(3), e12374. https://doi.org/10.1111/desc.12374

Cappagli, G., Finocchietti, S., Baud-Bovy, G., Badino, L., D'Ausilio, A., Cocchi, E., & Gori, M. (2018). Assessing social competence in visually impaired people and proposing an interventional program in visually impaired children. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 929-935. https://doi.org/10.1109/TCDS.2018.2809487

Cappagli, G., Finocchietti, S., Cocchi, E., Giammari, G., Zumiani, R., Cuppone, A. V., Baud-Bovy, G., & Gori, M. (2019). Audio motor training improves mobility and spatial cognition in visually impaired children. *Scientific Reports*, *9*(1), 3303. https://doi.org/10.1038/s41598-019-39981-x

Carlson, M., & Earls, F. (1997). Psychological and neuroendocrinological sequelae of early social deprivation in institutionalized children in Romania. In C. S. Carter, I. I. Lederhendler, & B. Kirkpatrick (Eds.), The integrative neurobiology of affiliation (pp. 419–428). New York Academy of Sciences.

Caron, V., Barras, A., van Nispen, R. M., & Ruffieux, N. (2023). Teaching social skills to children and adolescents with visual impairments: A systematic review. *Journal of Visual Impairment & Blindness*, *117*(2), 128-147. https://doi.org/0145482X231167150

Caron, V., Diacquenod, C., Melloul, V., & Ruffieux, N. (2021). Évolution des comportements sociaux d'enfants d'âge préscolaire avec déficience visuelle: Un projet pilote au jardin d'enfants inclusif du Centre pédagogique pour élèves handicapés de la vue (CPHV). *Revue suisse de pédagogie spécialisée*, *11*(1), 44–52-44–52. https://ojs.szh.ch/revue/article/view/147

Cascio, C. J., Moore, D., & McGlone, F. (2019). Social touch and human development. *Developmental Cognitive Neuroscience*, *35*, 5-11. https://doi.org/10.1016/j.dcn.2018.04.009

Cass, H. D., Sonksen, P. M., & McConachie, H. R. (1994). Developmental setback in severe visual impairment. *Archives of Disease in Childhood*, *70*(3), 192-196. https://doi.org/10.1136/adc.70.3.192

Castiello, U., Becchio, C., Zoia, S., Nelini, C., Sartori, L., Blason, L., D'Ottavio, G., Bulgheroni, M., & Gallese, V. (2010). Wired to be social: the ontogeny of human interaction. *PLoS One*, *5*(10), e13199. https://doi.org/10.1371/journal.pone.0013199

Cerritelli, F., Pizzolorusso, G., Ciardelli, F., La Mola, E., Cozzolino, V., Renzetti, C., D'Incecco, C., Fusilli, P., Sabatino, G., & Barlafante, G. (2013). Effect of osteopathic manipulative treatment on length of stay in a population of preterm infants: a randomized controlled trial. *BMC pediatrics*, *13*(1), 65. https://doi.org/10.1186/1471-2431-13-65

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive development*, *15*(4), 481-498. https://doi.org/10.1016/S0885-2014(01)00037-5

Chaumet-Riffaud, A. E., Chaumet-Riffaud, P., Cariou, A., Devisme, C., Audo, I., Sahel, J.-A., & Mohand-Said, S. (2017). Impact of retinitis pigmentosa on quality of life, mental health, and employment among young adults. *American journal of ophthalmology*, *177*, 169-174. https://doi.org/10.1016/j.ajo.2017.02.016

Chen, H., Liu, H., Lan, S., Wang, W., Qiao, Y., Li, Y., & Deng, G. (2025). DMAGaze: Gaze Estimation Based on Feature Disentanglement and Multi-Scale Attention. *arXiv preprint arXiv:2504.11160*. https://doi.org/10.48550/arXiv.2504.11160

Cowie, D., Makin, T. R., & Bremner, A. J. (2013). Children's responses to the rubber-hand illusion reveal dissociable pathways in body representation. *Psychological science*, *24*(5), 762-769. https://doi.org/10.1177/095679761246290

Dahl, A. (2017). Ecological commitments: Why developmental science needs naturalistic methods. *Child Development Perspectives*, *11*(2), 79-84. https://doi.org/10.1111/cdep.12217

Dale, N., & Salt, A. (2007). Early support developmental journal for children with visual impairment: the case for a new developmental framework for early intervention. *Child: care, health and development*, *33*(6), 684-690. https://doi.org/10.1111/j.1365-2214.2007.00798.x

Dale, N., & Sonksen, P. (2002). Developmental outcome, including setback, in young children with severe visual impairment. *Developmental medicine and child neurology*, *44*(9), 613-622. https://doi.org/10.1017/s0012162201002651

Dale, N. J., Tadić, V., & Sonksen, P. (2014). Social communicative variation in 1–3-year-olds with severe visual impairment. *Child: care, health and development*, *40*(2), 158-164. https://doi.org/10.1111/cch.12065

Damen, S., Janssen, M. J., Ruijssenaars, W. A., & Schuengel, C. (2015). Intersubjectivity effects of the high-quality communication intervention in people with deafblindness. *Journal of Deaf Studies and Deaf Education*, *20*(2), 191-201. https://doi.org/10.1093/deafed/env001

De Araújo, P. P., De Moura Filho, O. F., Valenti, V. E., Gallo, S. M., Camargo, M. R., Say, K. G., Marcomini, R. S., De Oliveira, G., Monteiro, C., & Wajnsztejn, R. (2014). Stabilometric parameters analysis in children with visual disorder. *International archives of medicine*, *7*(1), 1-6. https://doi.org/10.1186/1755-7682-7-1

DeCarlo, D. K., McGwin Jr, G., Bixler, M. L., Wallander, J., & Owsley, C. (2012). Impact of pediatric vision impairment on daily life: results of focus groups. *Optometry and Vision Science*, *89*(9), 1409-1416. https://doi.org/10.1097/OPX.0b013e318264f1dc

Della Longa, L., Dragovic, D., & Farroni, T. (2021). In touch with the heartbeat: newborns' cardiac sensitivity to affective and non-affective touch. *International journal of environmental research and public health*, *18*(5), 2212. https://doi.org/10.3390/ijerph18052212

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological review*, *3*(4), 357. https://doi.org/10.1037/h0070405

Dhawale, A. K., Smith, M. A., & Ölveczky, B. P. (2017). The role of variability in motor learning. *Annual Review of Neuroscience*, *40*, 479-498. https://doi.org/10.1146/annurev-neuro-072116-031548

Di Plinio, S., Scalabrini, A., & Ebisch, S. J. (2022). An integrative perspective on the role of touch in the development of intersubjectivity. *Brain and Cognition*, *163*, 105915. https://doi.org/10.1016/j.bandc.2022.105915

Diamond, L. M., & Fagundes, C. P. (2008). Developmental perspectives on links between attachment and affect regulation over the lifespan. In *Advances in child development and behavior* (Vol. 36, pp. 83-134). Elsevier. https://doi.org/10.1016/S0065-2407(08)00003-7

Donahue, S. P., Baker, C. N., Simon, G. R., Boudreau, A. D. A., Baker, C. N., Barden, G. A., Hackell, J. M., Hardin, A. P., Meade, K. E., & Moore, S. B. (2016). Procedures for the evaluation of the visual system by pediatricians. *Pediatrics*, *137*(1). https://doi.org/10.1542/peds.2015-3597

Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS One*, *8*(12), e82491. https://doi.org/10.1371/journal.pone.0082491

Dunbar, R. I., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*(5843), 1344-1347. https://doi.org/10.1126/science.1145463

Effenberg, A., Fehse, U., & Weber, A. (2011). Movement Sonification: Audiovisual benefits on motor learning. In BIO web of conferences (Vol. 1, p. 00022). EDP Sciences.

Effenberg, A. O., Fehse, U., Schmitz, G., Krueger, B., & Mechling, H. (2016). Movement sonification: Effects on motor learning beyond rhythmic adjustments. *Frontiers in Neuroscience*, *10*, 219. https://doi.org/10.3389/fnins.2016.00219

Elsman, E. B., Al Baaj, M., van Rens, G. H., Sijbrandi, W., van den Broek, E. G., van der Aa, H. P., Schakel, W., Heymans, M. W., de Vries, R., & Vervloed, M. P. (2019). Interventions to improve functioning, participation, and quality of life in children with visual impairment: a systematic review. *survey of ophthalmology*, *64*(4), 512-557. https://doi.org/10.1016/j.survophthal.2019.01.010

Ewaisha, M., El Shawarby, M., Abbas, H., & Sobh, I. (2020). End-to-end multitask learning for driver gaze and head pose estimation. *Electronic Imaging*, *32*, 1-6. https://doi.org/10.2352/ISSN.2470-1173.2020.16.AVM-110

Fairhurst, M. T., Löken, L., & Grossmann, T. (2014). Physiological and behavioral responses reveal 9-month-old infants' sensitivity to pleasant touch. *Psychological science*, *25*(5), 1124-1131. https://doi.org/10.1177/0956797614527114

Fang, Y., Nakashima, R., Matsumiya, K., Kuriki, I., & Shioiri, S. (2015). Eye-head coordination for visual cognitive processing. *PLoS One*, *10*(3), e0121035. https://doi.org/10.1371/journal.pone.0121035

Farroni, T., Johnson, M. H., & Csibra, G. (2004). Mechanisms of eye gaze perception during infancy. *Journal of cognitive neuroscience*, *16*(8), 1320-1326. https://doi.org/10.1162/0898929042304787

Feldman, R. (2012). Bio-behavioral synchrony: A model for integrating biological and microsocial behavioral processes in the study of parenting. *Parenting*, *12*(2-3), 154-164. https://doi.org/10.1080/15295192.2012.683342

Feldman, R., Eidelman, A. I., Sirota, L., & Weller, A. (2002). Comparison of skin-to-skin (kangaroo) and traditional care: parenting outcomes and preterm infant development. *Pediatrics*, *110*(1), 16-26. https://doi.org/10.1542/peds.110.1.16

Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., & Louzoun, Y. (2011). Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development*, *34*(4), 569-577. https://doi.org/10.1016/j.infbeh.2011.06.008

Feldman, R., Singer, M., & Zagoory, O. (2010). Touch attenuates infants' physiological reactivity to stress. *Developmental science*, *13*(2), 271-278. https://doi.org/10.1111/j.1467-7687.2009.00890.x

Ferber, S. G., Feldman, R., & Makhoul, I. R. (2008). The development of maternal touch across the first year of life. *Early human development*, *84*(6), 363-370. https://doi.org/10.1016/j.earlhumdev.2007.09.019

Ferrari, C., Vecchi, T., Merabet, L. B., & Cattaneo, Z. (2017). Blindness and social trust: The effect of early visual deprivation on judgments of trustworthiness. *Consciousness and cognition*, *55*, 156-164. https://doi.org/10.1016/j.concog.2017.08.005

Field, T. (2010). Touch for socioemotional and physical well-being: A review. *Developmental review*, *30*(4), 367-383. https://doi.org/10.1016/j.dr.2011.01.001

Foxx, R. M. (1977). Attention training: The use of overcorrection avoidance to increase the eye contact of autistic and retarded children. *Journal of Applied Behavior Analysis*, *10*(3), 489-499. https://doi.org/10.1901/jaba.1977.10-489

Fraiberg, S. (1968). Parallel and divergent patterns in blind and sighted infants. *The Psychoanalytic study of the child*, *23*(1), 264-300. https://doi.org/10.1080/00797308.1968.11822959

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, *97*(458), 611-631. https://doi.org/10.1198/016214502760047131

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in ecology and evolution*, *7*(11), 1325-1330. https://doi.org/10.1111/2041-210X.12584
Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16), R724-R732. https://doi.org/10.1016/j.cub.2007.05.068
Gallace, A., & Spence, C. (2016). Social touch. In *Affective touch and the neurophysiology of CT afferents* (pp. 227-238). Springer. https://doi.org/10.1007/978-1-4939-6418-5_14
Gallese, V., & Ebisch, S. (2013). Embodied simulation and touch: The sense of touch in social cognition. *Phenomenology and Mind*(4), 196-210. https://doi.org/10.13128/Phe_Mi-19602
Gao, Z., Chen, J., Liu, Y., Jin, Y., & Tian, D. (2025). A systematic survey on human pose estimation: upstream and downstream tasks, approaches, lightweight models, and prospects. *Artificial Intelligence Review*, *58*(3), 68. https://doi.org/10.1007/s10462-024-11060-2
Gentsch, A., Panagiotopoulou, E., & Fotopoulou, A. (2015). Active interpersonal touch gives rise to the social softness illusion. *Current Biology*, *25*(18), 2392-2397. https://doi.org/10.1016/j.cub.2015.07.049
Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. Oxford University Press, USA.
Gibson, E. J., & Schumuckler, M. A. (1989). Going somewhere: An ecological and experimental approach to development of mobility. *Ecological psychology*, *1*(1), 3-25. https://doi.org/10.1207/s15326969eco0101_2
Gibson, J. J. (1979). The ecological approach to visual perception. Boston, MA: Houghton Mifflin.
Gilbert, M. (2007). Mutual recognition, common knowledge, and joint attention. *Hommage à Wlodek: Philosophical papers dedicated to Wlodek Rabinowicz*, 1-21.
Gima, H., Kihara, H., Watanabe, H., Nakano, H., Nakano, J., Konishi, Y., Nakamura, T., & Taga, G. (2018). Early motor signs of autism spectrum disorder in spontaneous position and movement of the head. *Experimental brain research*, *236*(4), 1139-1148. https://doi.org/10.1007/s00221-018-5202-x
Gori, M. (2015). Multisensory integration and calibration in children and adults with and without sensory and motor disabilities. *Multisensory Research*, *28*(1-2), 71-99. https://doi.org/10.1163/22134808-00002478
Gori, M., Cappagli, G., Tonelli, A., Baud-Bovy, G., & Finocchietti, S. (2016). Devices for visually impaired people: High technological devices with low user acceptance and no adaptability for children. *Neuroscience & Biobehavioral Reviews*, *69*, 79-88. https://doi.org/10.1016/j.concog.2017.08.005
Graham, K. E., Buryn-Weitzel, J. C., Lahiff, N. J., Wilke, C., & Slocombe, K. E. (2021). Detecting joint attention events in mother-infant dyads: Sharing looks cannot be reliably identified by naïve third-party observers. *PLoS One*, *16*(7), e0255241. https://doi.org/10.1371/journal.pone.0255241
Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*, 1167. https://doi.org/10.3389/fpsyg.2015.01167
Green, E., Stroud, L., Bloomfield, S., Cronje, J., Foxcroft, C., Hurter, K., Lane, H., Marais, R., Marx, C., & McAlinden, P. (2016). Griffiths scales of child development. *Hogrefe Ltd.: Oxford, UK*.
Gross, J. J. (2013). *Handbook of emotion regulation*. Guilford publications.
Grumi, S., Cappagli, G., Aprile, G., Mascherpa, E., Gori, M., Provenzi, L., & Signorini, S. (2021). Togetherness, beyond the eyes: a systematic review on the interaction between visually impaired children and their parents. *Infant Behavior and Development*, *64*, 101590. https://doi.org/10.1016/j.infbeh.2021.101590

Guarischi, M., Hu, Y., Kurt, A. B., Zanchi, S., Barth, E., & Gori, M. (2024). A Machine Learning Approach to Unveil Balance Behavior Through Aging with an Auditory Cue. 2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA),

Guarischi, M., Montagnani, E., Catalano, G., Saligari, E., Signorini, S., & Gori, M. (2025). From motion to interaction: How multisensory information shapes motor behaviors in children with visual impairment. *Research in developmental disabilities*, *159*, 104956. https://doi.org/10.1016/j.ridd.2025.104956

Gui, A., Perelli, D., Rizzo, G., Ferruzza, E., & Mercuriali, E. (2023). Children’s total blindness as a risk factor for early parent-child relationships: Preliminary findings from an Italian sample. *Frontiers in Psychology*, *14*, 1175675. https://doi.org/10.3389/fpsyg.2023.1175675

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature neuroscience*, *5*(4), 382-385. https://doi.org/10.1038/nn827

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental brain research*, *126*(1), 128-133. https://doi.org/10.1007/s002210050722

Hammal, Z., & Cohn, J. F. (2014). Intra-and interpersonal functions of head motion in emotion communication. In Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges (pp. 19-22).

Hammal, Z., Cohn, J. F., Messinger, D. S., Mattson, W. I., & Mahoor, M. H. (2013). Head movement dynamics during normal and perturbed parent-infant interaction. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (pp. 276-282). IEEE.

Han, J., Waddington, G., Adams, R., Anson, J., & Liu, Y. (2016). Assessing proprioception: a critical review of methods. *Journal of sport and health science*, *5*(1), 80-90. https://doi.org/10.1016/j.jshs.2014.10.004

Hanna, J. E., Brennan, S. E., & Savietta, K. J. (2020). Eye gaze and head orientation cues in face-to-face referential communication. *Discourse Processes*, *57*(3), 201-223. https://doi.org/10.1080/0163853X.2019.1675467

Harlow, H. F., & Zimmermann, R. R. (1958). The development of affectional responses in infant monkeys. *Proceedings of the American Philosophical Society*, *102*(5), 501-509.

Heidegger, M. (1927). Sein und Zeit. Tübingen: Niemeyer. *Original work published*.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, *56*(5), 872. https://doi.org/10.1037/h0040546

Hermann, T., Höner, O., & Ritter, H. (2005). AcouMotion–an interactive sonification system for acoustic motion control. In: Gibet, S., Courty, N., Kamp, JF. (eds) Gesture in Human-Computer Interaction and Simulation. GW 2005. Lecture Notes in Computer Science(), vol 3881. Springer, Berlin, Heidelberg.

Herrera, E., Reissland, N., & Shepherd, J. (2004). Maternal touch and maternal child-directed speech: effects of depressed mood in the postnatal period. *Journal of affective disorders*, *81*(1), 29-39. https://doi.org/10.1016/j.jad.2003.07.001

Hertenstein, M. J. (2002). Touch: Its communicative functions in infancy. *Human Development*, *45*(2), 70-94. https://doi.org/10.1159/000048154

Herzfeld, D. J., & Shadmehr, R. (2014). Motor variability is not noise, but grist for the learning mill. *Nature neuroscience*, *17*(2), 149-150. https://doi.org/10.1038/nn.3633

Hobson, R. P. (2025). Intersubjectivity: A Unit of Analysis for Developmental Psychology. In J. Delafield Butt & V. Reddy (Eds.), *Intersubjective Minds: Rhythm, Sympathy, and Human Being* (pp. 0). Oxford University Press. https://doi.org/10.1093/oso/9780192865373.003.0007

Hofer, M. A. (1994). Hidden regulators in attachment, separation, and loss. *Monographs of the Society for Research in Child Development*, 192-207. https://doi.org/10.2307/1166146

Husserl, E. (1952). Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie: drittes Buch, Die Phänomenologie und die Fundamente der Wissenschaften. *Husserliana.*

Hyvärinen, L., Näsänen, R., & Laurinen, P. (1980). New visual acuity test for pre-school children. *Acta ophthalmologica*, *58*(4), 507-511. https://doi.org/10.1111/j.1755-3768.1980.tb08291.x

Jean, A. D., & Stack, D. M. (2012). Full-term and very-low-birth-weight preterm infants' self-regulating behaviors during a Still-Face interaction: influences of maternal touch. *Infant Behavior and Development*, *35*(4), 779-791. https://doi.org/10.1016/j.infbeh.2012.07.023

Jonas, H. (1973). *Organismus und Freiheit: Ansätze zu einer philosophischen Biologie*. Vandenhoeck & Ruprecht.

Jongerius, C., Hessels, R. S., Romijn, J. A., Smets, E. M., & Hillen, M. A. (2020). The measurement of eye contact in human interactions: a scoping review. *Journal of Nonverbal Behavior*, *44*, 363-389. https://doi.org/10.1007/s10919-020-00333-3

Kallie, C. S., Schrater, P. R., & Legge, G. E. (2007). Variability in stepping direction explains the veering behavior of blind walkers. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 183. https://doi.org/10.1037/0096-1523.33.1.183

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual review of psychology*, *62*(1), 1-21. https://doi.org/10.1146/annurev.psych.031809.130730

Kida, T., & Shinohara, K. (2013). Gentle touch activates the prefrontal cortex in infancy: an NIRS study. *Neuroscience letters*, *541*, 63-66. https://doi.org/10.1016/j.neulet.2013.01.048

Klein, M. (1957). Envy and Gratitude *Reprinted in The Writings of Melanie Klein*, *3*, 176-235.

Lamb, M. E., & Lerner, R. M. (2015). *Handbook of child psychology and developmental science: Socioemotional processes, Vol. 3*. John Wiley & Sons, Inc.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision research*, *41*(25-26), 3559-3565. https://doi.org/10.1016/S0042-6989(01)00102-X

Langton, S. R., & Bruce, V. (2000). You must see the point: automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 747.

Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: the 2-month transition. *Developmental psychology*, *41*(1), 265. https://doi.org/10.1037/0012-1649.41.1.265

Lee, D. K., Cole, W. G., Golenia, L., & Adolph, K. E. (2018). The cost of simplifying complex developmental phenomena: A new perspective on learning to walk. *Developmental science*, *21*(4), e12615. https://doi.org/10.1111/desc.12615

Lerner, R. M., Liben, L. S., & Mueller, U. (2015). *Handbook of child psychology and developmental science, cognitive processes* (Vol. 2). John Wiley & Sons.

Lew-Williams, C., Ferguson, B., Abu-Zhaya, R., & Seidl, A. (2019). Social touch interacts with infants' learning of auditory patterns. *Developmental Cognitive Neuroscience*, *35*, 66-74. https://doi.org/10.1016/j.dcn.2017.09.006

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental psychobiology*, *56*(2), 292-315. https://doi.org/10.1002/dev.21197

Loggia, M. L., Juneau, M., & Bushnell, M. C. (2011). Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. *PAIN®*, *152*(3), 592-598. https://doi.org/10.1016/j.pain.2010.11.032

Löken, L. S., Wessberg, J., Morrison, I., McGlone, F., & Olausson, H. (2009). Coding of pleasant touch by unmyelinated afferents in humans. *Nature neuroscience*, *12*(5), 547-548. https://doi.org/10.1038/nn.2312

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, *37*(9), 1443-1457. https://doi.org/10.1068/p5896

Lueck, A. H. (2008). *Developmental guidelines for infants with visual impairments: A guidebook for early intervention*. American Printing House for the Blind, Incorporated.

Main, M. (2000). The organized categories of infant, child, and adult attachment: Flexible vs. inflexible attention under attachment-related stress. *Journal of the American Psychoanalytic Association*, *48*(4), 1055-1096. https://doi.org/10.1177/00030651000480041801

Maitre, N. L., Stark, A. R., Menser, C. C. M., Chorna, O. D., France, D. J., Key, A. F., Wilkens, K., Moore-Clingenpeel, M., Wilkes, D. M., & Bruehl, S. (2017). Cry presence and amplitude do not reflect cortical processing of painful stimuli in newborns with distinct responses to touch or cold. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *102*(5), F428-F433.

Manzotti, A., Cerritelli, F., Esteves, J. E., Lista, G., Lombardi, E., La Rocca, S., Gallace, A., McGlone, F. P., & Walker, S. C. (2019). Dynamic touch reduces physiological arousal in preterm infants: A role for c-tactile afferents? *Developmental Cognitive Neuroscience*, *39*, 100703. https://doi.org/10.1016/j.dcn.2019.100703

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and psychopathology*, *22*(3), 491-495. https://doi.org/10.1017/S0954579410000222

Maurer, D., & Maurer, C. (1988). *The world of the newborn*. New York: Basic Books.

McGlone, F., Wessberg, J., & Olausson, H. (2014). Discriminative and affective touch: sensing and feeling. *Neuron*, *82*(4), 737-755. https://doi.org/10.1016/j.neuron.2014.05.001 External Link

McKay, M. J., Baldwin, J. N., Ferreira, P., Simic, M., Vanicek, N., Wojciechowski, E., Mudge, A., Burns, J., & Consortium, N. P. (2017). Spatiotemporal and plantar pressure patterns of 1000 healthy individuals aged 3–101 years. *Gait & posture*, *58*, 78-87. https://doi.org/10.1016/j.gaitpost.2017.07.004

Meltzoff, A. N., & Brooks, R. (2001). "Like Me" as a Building Block for Understanding Other Minds: Bodily Acts. *Intentions and intentionality: Foundations of social cognition*, 171.

Merleau-Ponty, M. (1968). *The visible and the invisible: Followed by working notes*. Northwestern University Press.

Merleau-Ponty, M., Landes, D., Carman, T., & Lefort, C. (2013). *Phenomenology of perception*. Routledge.

Merleu-Ponty, M. (1962). Phenomenology of perception. *Trans. Colin Smith. New York: Humanities Press, and London: Routledge & Kegan Paul*.

Moll, H., & Meltzoff, A. N. (2011). 15 Joint Attention as the Fundamental Basis of Understanding Perspectives. *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*, *393*.

Montagnani, E., Bradley, H., & Smith, B. A. (2024). Gait Characteristics and Development in Pediatric Populations With Visual Disorders: Where Do We Stand and Where Are We Going? *Journal of Visual Impairment & Blindness*. https://doi.org/10.1177/0145482X241255961

Montagu, A. (1971). Touching: The human significance of the skin. New York, NY: Columbia University Press.

Morales, M., Mundy, P., Delgado, C. E., Yale, M., Neal, R., & Schwartz, H. K. (2000). Gaze following, temperament, and language development in 6-month-olds: A replication and extension. *Infant Behavior and Development*, *23*(2), 231-236. https://doi.org/10.1016/S0163-6383(01)00038-8

Morrison, I., Löken, L. S., & Olausson, H. (2010). The skin as a social organ. *Experimental brain research*, *204*(3), 305-314. https://doi.org/10.1007/s00221-009-2007-y

Murphy-Chutorian, E., & Trivedi, M. M. (2008). Head pose estimation in computer vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *31*(4), 607-626. https://doi.org/10.1109/TPAMI.2008.106

Nardini, M., Cowie, D., Bremner, A., Lewkowicz, D., & Spence, C. (2012). The development of multisensory balance, locomotion, orientation, and navigation. *Multisensory development*, 137-158.

Newell, K. M., & Corcos, D. M. (1993). Variability and motor control. Champaign, IL: Human Kinetics Publishers.

Noë, A. (2004). *Action in perception*. MIT press.

Nown, T. H., Upadhyay, P., Kerr, A., Andonovic, I., Tachtatzis, C., & Grealy, M. A. (2022). A mapping review of real-time movement sonification systems for movement rehabilitation. *IEEE Reviews in Biomedical Engineering*, *16*, 672-686. https://doi.org/10.1109/RBME.2022.3187840

Pallini, S., Morelli, M., Chirumbolo, A., Baiocco, R., Laghi, F., & Eisenberg, N. (2019). Attachment and attention problems: A meta-analysis. *Clinical psychology review*, *74*, 101772. https://doi.org/10.1016/j.cpr.2019.101772

Pataky, T. C. (2012). One-dimensional statistical parametric mapping in Python. *Computer methods in biomechanics and biomedical engineering*, *15*(3), 295-301. https://doi.org/10.1080/10255842.2010.527837

Piaget, J., & Cook, M. (1952). *The origins of intelligence in children* (Vol. 8). International universities press New York.

Pirazzoli, L., Lloyd-Fox, S., Braukmann, R., Johnson, M. H., & Gliga, T. (2019). Hand or spoon? Exploring the neural basis of affective touch in 5-month-old infants. *Developmental Cognitive Neuroscience*, *35*, 28-35. https://doi.org/10.1016/j.dcn.2018.06.002

Pizzera, A., & Hohmann, T. (2015). Acoustic information during motor control and action perception: a review. The Open Psychology Journal, 8(1). https://doi.org/10.2174/1874350101508010183

Price, C., Montagnani, E., Santos, A. M., Nester, C., & Morrison, S. (2022). Longitudinal study of foot pressures during real-world walking as infants develop from new to confident walkers. *Gait & posture*, *92*, 351-358. https://doi.org/10.1016/j.gaitpost.2021.12.003

Pring, L. E. (2005). *Autism and blindness: Research and reflections*. Whurr Publishers.

Provenzi, L., Pettenati, G., Luparia, A., Paini, D., Aprile, G., Morelli, F., Mascherpa, E., Vercellino, L., Grumi, S., & Signorini, S. (2022). Case report: Dancing in the dark: A critical single case study engaging a blind father in the rehabilitation journey of his visually impaired child. *Frontiers in Psychology*, *13*, 942321. https://doi.org/10.3389/fpsyg.2022.942321

Rainey, L., Elsman, E. B. M., van Nispen, R. M. A., van Leeuwen, L. M., & van Rens, G. H. M. B. (2016). Comprehending the impact of low vision on the lives of children and adolescents: a qualitative approach. *Quality of Life Research*, *25*, 2633-2643. https://doi.org/10.1007/s11136-016-1292-8

Ranganathan, R., Lee, M.-H., & Newell, K. M. (2020). Repetition without repetition: Challenges in understanding behavioral flexibility in motor skill. *Frontiers in Psychology*, *11*, 2018. https://doi.org/10.3389/fpsyg.2020.02018

Ratcliffe, M. (2013). Touch and the sense of reality. *The hand, an organ of the mind: What the manual tells the mental*, *131*.
Reynell, J., & Zinkin, P. (1975). New procedures for the developmental assessment of young children with severe visual handicaps. *Child: care, health and development*, *1*(1), 61-69. https://doi.org/10.1111/j.1365-2214.1975.tb00203.x
Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in neurosciences*, *21*(5), 188-194. https://doi.org/10.1016/S0166-2236(98)01260-0
Rochat, P. (2014). *Early social cognition: Understanding others in the first months of life*. Psychology Press.
Rodger, M. W., & Craig, C. M. (2016). Beyond the metronome: auditory events and music may afford more than just interval durations as gait cues in Parkinson's disease. *Frontiers in Neuroscience*, *10*, 272. https://doi.org/10.3389/fnins.2016.00272
Rothbart, M. K., Posner, M. I., & Kieras, J. (2006). Temperament, attention, and the development of self-regulation. https://doi.org/10.1002/9780470757703.ch17
Ruiz, N., Chong, E., & Rehg, J. M. (2018). Fine-grained head pose estimation without keypoints. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 2074-2083).
Sakurai, T. (2022). Social processes and social environment during development. In Seminars in Cell & Developmental Biology (Vol. 129, pp. 40-46). Academic Press.
Salminen, A.-L., & Karhula, M. E. (2014). Young persons with visual impairment: Challenges of participation. *Scandinavian journal of occupational therapy*, *21*(4), 267-276. https://doi.org/10.3109/11038128.2014.899622
Schiatti, L., Cappagli, G., Martolini, C., Maviglia, A., Signorini, S., Gori, M., & Crepaldi, M. (2020). *A Novel Wearable and Wireless Device to Investigate Perception in Interactive Scenarios* (Vol. 2020). https://doi.org/10.1109/EMBC44109.2020.9176167
Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, *5*(7), 532-545. https://doi.org/10.1038/nrn1427
Shai, D., & Belsky, J. (2011). Parental embodied mentalizing: Let's be explicit about what we mean by implicit. *Child Development Perspectives*, *5*(3), 187-188. https://doi.org/10.1111/j.1750-8606.2011.00195.x
Sherrington, C. S. (1913). Reflex inhibition as a factor in the co-ordination of movements and postures. *Quarterly Journal of Experimental Physiology: Translation and Integration*, *6*(3), 251-310. https://doi.org/10.1113/expphysiol.1913.sp000142
Simpson, E. A., Sclafani, V., Paukner, A., Kaburu, S. S., Suomi, S. J., & Ferrari, P. F. (2019). Handling newborn monkeys alters later exploratory, cognitive, and social behaviors. *Developmental Cognitive Neuroscience*, *35*, 12-19. https://doi.org/10.1016/j.dcn.2017.07.010
Sroufe, L. A. (1977). Wariness of strangers and the study of infant development. *Child development*, 731-746. https://doi.org/10.2307/1128323
Stack, D. M., & Muir, D. W. (1990). Tactile stimulation as a component of social interchange: New interpretations for the still-face effect. *British Journal of Developmental Psychology*, *8*(2), 131-145. https://doi.org/10.1111/j.2044-835X.1990.tb00828.x
Steenson, C. J., & Rodger, M. W. (2015). Bringing sounds into use: thinking of sounds as materials and a sketch of auditory affordances. *Open Psychology Journal*, *8*(Suppl 3), 174-182. https://doi.org/10.2174/1874350101508010174
Stern, D. N. (1971). A micro-analysis of mother-infant interaction. behavior regulating social contact between a mother and her 3 1/2 Month-Old Twins. *Journal of the american academy of child psychiatry*, *10*(3), 501-517.
Stiefelhagen, R., & Zhu, J. (2002). Head orientation and gaze direction in meetings. In CHI'02 Extended Abstracts on Human Factors in Computing Systems (pp. 858-859).

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in cognitive sciences*, *10*(10), 471-476. https://doi.org/10.1016/j.tics.2006.08.006 External Link

Tadić, V., Pring, L., & Dale, N. (2009). Attentional processes in young children with congenital visual impairment. *British Journal of Developmental Psychology*, *27*(2), 311-330. https://doi.org/10.1348/026151008x310210

Tamis-LeMonda, C. S., Adolph, K. E., Lobo, S. A., Karasik, L. B., Ishak, S., & Dimitropoulou, K. A. (2008). When infants take mothers' advice: 18-month-olds integrate perceptual and social information to guide motor action. *Developmental psychology*, *44*(3), 734. https://doi.org/10.1037/0012-1649.44.3.734

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental science*, *20*(6), e12456. https://doi.org/10.1111/desc.12456

Teller, D. Y., McDonald, M. A., Preston, K., Sebris, S. L., & Dobson, V. (1986). Assessment of visual acuity in infants and children; the acuity card procedure. *Developmental Medicine & Child Neurology*, *28*(6), 779-789. https://doi.org/10.1111/j.1469-8749.1986.tb03932.x

Thelen, E. (2005). Dynamic systems theory and the complexity of change. *Psychoanalytic dialogues*, *15*(2), 255-283. https://doi.org/10.1080/10481881509348831

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT press.

Tomasello, M. (2014). Joint attention as social cognition. In *Joint attention* (pp. 103-130). Psychology Press.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental science*, *10*(1), 121-125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675-691. https://doi.org/10.1017/S0140525X05000129

Tomenotti, F. F., Noceti, N., & Odone, F. (2024). Head pose estimation with uncertainty and an application to dyadic interaction detection. *Computer Vision and Image Understanding*, *243*, 103999. https://doi.org/10.1016/j.cviu.2024.103999

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. *Before speech: The beginning of interpersonal communication*, *1*, 530-571.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. *Intersubjective communication and emotion in early ontogeny*, *15*, 46.

Tronick, E. Z., & Cohn, J. F. (1989). Infant-mother face-to-face interaction: Age and gender differences in coordination and the occurrence of miscoordination. *Child development*, 85-92. https://doi.org/10.2307/1131074

Tuthill, J. C., & Azim, E. (2018). Proprioception. *Current Biology*, *28*(5), R194-R203. https://doi.org/10.1016/j.cub.2018.01.064

Tuulari, J. J., Scheinin, N. M., Lehtola, S., Merisaari, H., Saunavaara, J., Parkkola, R., Sehlstedt, I., Karlsson, L., Karlsson, H., & Björnsdotter, M. (2019). Neural correlates of gentle skin stroking in early infancy. *Developmental Cognitive Neuroscience*, *35*, 36-41. https://doi.org/10.1016/j.dcn.2017.10.004

Urqueta Alfaro, A., Vacaru, S., Wittich, W., & Sterkenburg, P. S. (2021). Identifying when children with visual impairment share attention: A novel protocol and the impact of visual acuity. *Infant Behavior and Development*, *64*, 101585. https://doi.org/10.1016/j.infbeh.2021.101585

Væver, M., Beebe, B., Kirk, O., Snidmann, N., Harder, S., & Tronick, E. (2015). An automated approach for measuring infant head orientation in a face-to-face interaction. *Behavior Research Methods*, *47*, 328-339. https://doi.org/10.3758/s13428-014-0487-6

Vallbo, Å., Olausson, H., & Wessberg, J. (1999). Unmyelinated afferents constitute a second system coding tactile stimuli of the human hairy skin. *Journal of Neurophysiology*, *81*(6), 2753-2763. https://doi.org/10.1152/jn.1999.81.6.275

van den Broek, E. G., van Eijden, A. J., Overbeek, M. M., Kef, S., Sterkenburg, P. S., & Schuengel, C. (2017a). A systematic review of the literature on parenting of young children with visual impairments and the adaptions for video-feedback intervention to promote positive parenting (VIPP). *Journal of Developmental and Physical Disabilities*, *29*, 503-545. https://doi.org/10.1007/s10882-016-9529-6

van den Broek, E. G., van Eijden, A. J., Overbeek, M. M., Kef, S., Sterkenburg, P. S., & Schuengel, C. (2017b). A systematic review of the literature on parenting of young children with visual impairments and the adaptions for video-feedback intervention to promote positive parenting (VIPP). *Journal of Developmental and Physical Disabilities*, *29*(3), 503-545.

Van Hecke, A. V., Mundy, P., Block, J. J., Delgado, C. E., Parlade, M. V., Pomares, Y. B., & Hobson, J. A. (2012). Infant responding to joint attention, executive processes, and self-regulation in preschool children. *Infant Behavior and Development*, *35*(2), 303-311. https://doi.org/10.1016/j.infbeh.2011.12.001

Verver, S. H., Vervloed, M. P., & Steenbergen, B. (2019). The use of augmented toys to facilitate play in school-aged children with visual impairments. *Research in developmental disabilities*, *85*, 70-81. https://doi.org/10.1016/j.ridd.2018.11.006

Verver, S. H., Vervloed, M. P., & Steenbergen, B. (2020). Characteristics of peer play in children with visual impairments. *Research in developmental disabilities*, *105*, 103714. https://doi.org/10.1016/j.ridd.2020.103714

Vervloed, M. P., van den Broek, E. C., & van Eijden, A. J. (2020). Critical review of setback in development in young children with congenital blindness or visual impairment. *International Journal of Disability, Development and Education*, *67*(3), 336-355. https://doi.org/10.1080/1034912X.2019.1588231

Von Hofsten, C. (2009). Action, the foundation for cognitive development. *Scandinavian Journal of Psychology*, *50(6)*, 617-623. https://doi.org/10.1111/j.1467-9450.2009.00780.x

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.

Waibel, A., Schultz, T., Bett, M., Denecke, M., Malkin, R., Rogina, I., Stiefelhagen, R., & Yang, J. (2003). SMaRT: The smart meeting room task at ISL. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing. In 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). (Vol. 4, pp. IV-752). IEEE,

Walicka-Cupryś, K., Rachwał, M., Guzik, A., & Piwoński, P. (2022). Body balance of children and youths with visual impairment (pilot study). *International journal of environmental research and public health*, *19*(17), 11095. https://doi.org/10.3390/ijerph191711095

Walker, S. C., Trotter, P. D., Swaney, W. T., Marshall, A., & Mcglone, F. P. (2017). C-tactile afferents: Cutaneous mediators of oxytocin release during affiliative tactile interactions? *Neuropeptides*, *64*, 27-38. https://doi.org/10.1016/j.npep.2017.01.001

Weinberg, M. K., & Tronick, E. Z. (1994). Beyond the face: An empirical study of infant affective configurations of facial, vocal, gestural, and regulatory behaviors. *Child development*, *65*(5), 1503-1515. https://doi.org/10.1111/j.1467-8624.1994.tb00832.x

Winnicott, D. W. (1948). Pediatrics and psychiatry. *British Journal of Medical Psychology*, *21*(4), 229-240. https://doi.org/10.1111/j.2044-8341.1948.tb01173.x

Winnicott, D. W. (1964). *The child, the family, and the outside world*. Penguin UK.

Winter, D. A., Sidwall, H. G., & Hobson, D. A. (1974). Measurement and reduction of noise in kinematics of locomotion. *Journal of biomechanics*, *7*(2), 157-159. https://doi.org/10.1016/0021-9290(74)90056-6

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12), 739-751. https://doi.org/10.1038/nrn3112

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, *17*(2), 89-100.

You, J., Li, Z., & Du, J. (2023). A new iterative initialization of EM algorithm for Gaussian mixture models. *PLoS One*, *18*(4), e0284114. https://doi.org/10.1371/journal.pone.0284114

Zhao, Z., Zhu, Z., Zhang, X., Tang, H., Xing, J., Hu, X., Lu, J., & Qu, X. (2022). Identifying autism with head movement features by implementing machine learning algorithms. *Journal of Autism and Developmental Disorders*, *52*(7), 3038-3049. https://doi.org/10.1007/s10803-021-05179-2