InteroSense: Earable Predictive Sensory Intelligence for Human State-Aware Robot Adaptation in Autism Support

Maha M. A. Al-Obaidli, John-John Cabibihan
Department of Mechanical and Industrial Engineering
Qatar University
Doha, Qatar

***Abstract*—Socially assistive robots can support autistic children through structured, repeatable, and engaging interaction. Yet many adaptive systems still react after visible distress, disengagement, refusal, or challenging behavior has already emerged. InteroSense reframes robot-assisted autism support as a predictive human state-aware robotics problem: can an earable sensing layer, synchronized with robot and session context, estimate child-specific sensory-state trajectories early enough to support therapist-guided adaptation? The proposed framework treats the earable as a sensory gateway rather than a simple monitor. It combines candidate physiological, head-motion, temperature, and privacy-preserving auditory-exposure features with robot behavior logs, task demand, transition timing, environmental context, and therapist annotations. These streams update a personalized sensory-state profile and an adaptive sensory digital twin that estimates states such as regulated engagement, rising sensory load, overload risk, disengagement risk, and recovery with confidence and explanation. The core advance is the end-to-end pathway from multimodal early indicators to bounded robot, task, session, and caregiver-reporting decisions. The paper presents the architecture, candidate indicators, modeling direction, evaluation plan, and ethical guardrails for work-in-progress research at the intersection of earable sensing, predictive HRI, adaptive assistive robotics, and child-centered AI.**

***Index Terms*—human state-aware robotics; earable sensing; predictive sensory intelligence; autism support; adaptive social robots; sensory overload; digital twin; ethical HRI**

I. Introduction

Socially assistive robots (SARs) have been widely explored for autism support because they can provide predictable routines, repeatable prompting, and engaging learning interactions [1], [2]. In practice, however, the robot is not only an instructional agent. It is also a sensory event: its voice, motion, proximity, timing, lights, transitions, and prompt density can either support regulation or increase sensory load. This matters because sensory sensitivity and sensory-processing differences are central to many autistic children's everyday experiences, and the same robot behavior may be supportive for one child but overwhelming for another.

A central limitation in many adaptive HRI systems is that adaptation is triggered after behavior becomes visible: gaze aversion, refusal, agitation, shutdown, or challenging behavior. Therapists and caregivers often need a different form of support: earlier evidence that a child is moving away from a manageable sensory state, before escalation becomes behaviorally obvious. Recent wearable work in autism has shown the relevance of physiological and movement features for detecting or forecasting challenging behavior [4]–[6]; yet prediction alone is not sufficient for HRI. A robot system must also decide how to convert uncertain human-state estimates into safe, interpretable, and human-supervised adaptation.

InteroSense addresses this design gap by proposing a predictive sensory-intelligence layer for autism-support robotics. The framework links four components: (1) an earable sensory gateway; (2) synchronized robot/session context; (3) a personalized sensory-state profile and adaptive sensory digital twin; and (4) therapist-guided prediction-to-adaptation policies. The aim is not to claim diagnosis, emotion recognition, or independent clinical judgment. The aim is to provide earlier, explainable decision support for pacing, modality selection, transition design, recovery windows, and caregiver communication.

At a broader HRI level, the framework examines how earable-plus-context sensory-state estimation can support trustworthy, context-sensitive robot adaptation: uncertain multimodal outputs remain interpretable, calibrated, and constrained by therapist oversight [11], [12]. In Qatar-based deployment contexts, this motivates attention to Arabic-language interaction, family-mediated care, privacy expectations around child biosignals and auditory context, and culturally appropriate therapist-robot-caregiver coordination [13], [14].

Work-in-progress contribution. This paper contributes: (C1) an earable-to-robot framework for early sensory-state estimation in autism-support HRI; (C2) a prediction-to-adaptation loop that fuses physiological, motion, auditory-context, robot-behavior, and session-context features; (C3) therapist-supervised policies for bounded robot, task, and session adaptation; and (C4) a staged validation pathway covering bench testing, expert review, ethics approval, child-facing feasibility, predictive assessment, and failure-mode safeguards.

II. Research Basis and Design Gap

Evidence on robot-mediated autism interventions suggests that robots can support learning and social engagement, but long-term, real-world, and clinically grounded personalization remain difficult [1], [2]. Work on sensory-focused robotics has also recognized that robot-assisted interaction can be used to structure sensory experiences for autistic children [3]. In parallel, wearable-sensing studies have reported that physiological features, including heart-rate and HRV-related measures, can be informative during robot or classroom interactions involving autistic children [4], [5]. A recent classroom study further suggests that multimodal wearable

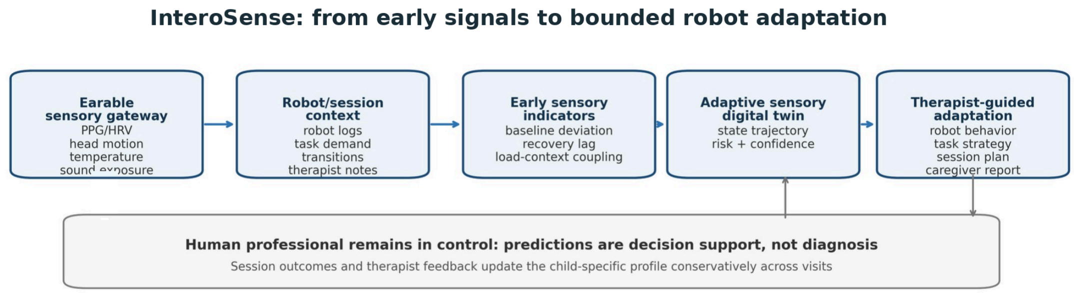

Fig. 1. InteroSense as an earable-to-robot predictive sensory-intelligence loop. Earable streams are interpreted together with robot/session context, converted into child-specific early indicators, updated through an adaptive sensory digital twin, and translated into therapist-guided support across robot behavior, task strategy, session planning, and caregiver reporting.

time-series may provide lead time for proactive support in real-world educational settings [5].

Earables add a strategically meaningful sensing site to this landscape. Prior work has shown the feasibility of in-ear physiological sensing [7], ear-mounted PPG for stress-related classification [8], and the rapid growth of earable systems for health, contextual awareness, and interaction [9]. InteroSense does not assume that the ear is always superior to wrist, chest, camera, or room sensing. Instead, it treats the earable as a promising research site for supervised short sessions because it can combine autonomic signals, head movement, temperature trends, and auditory-context exposure near the child's listening environment.

The remaining gap is therefore specific: existing autism-support robotics can structure interaction, and wearable sensing can support detection or forecasting, but limited work closes the loop from earable-based early sensory indicators to bounded, therapist-guided robot and session-level adaptation. Table I summarizes the prediction-to-adaptation logic that connects sensing, state estimation, personalization, and therapist-supervised support.

III. Proposed Framework

A. Earable Sensory Gateway

The earable is designed as a child-centered sensory gateway for supervised sessions. Candidate form factors include soft behind-ear, ear-hook, or ear-adjacent prototypes that prioritize comfort, safe removal, hygiene, and tolerance before long-duration performance. Candidate signals include PPG-derived heart-rate dynamics and HRV features, temperature trends, inertial head-motion and orientation features, and privacy-preserving auditory-exposure descriptors such as intensity, variability, and coarse acoustic load. Raw speech should not be stored by default.

What makes the approach technically meaningful is not the sensor location alone. Each sensor window is synchronized with robot behavior and session metadata: activity type, prompt frequency, speech rate, turn-taking latency, movement amplitude, proximity, transition timing, visual stimulation level, task difficulty, and therapist notes. This allows the system to ask a richer question: not only whether regulation is changing,

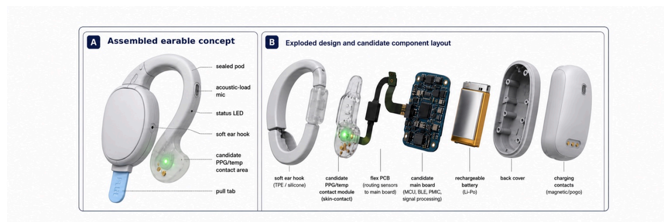

Fig. 2. Conceptual InteroSense earable form factor and candidate component layout. The figure represents design intent for WIP validation; comfort, sensor placement, thermal safety, motion artifacts, hygiene, and signal quality must be verified through bench, adult, and expert-review testing before child-facing studies.

but which robot, task, or environmental conditions may be contributing to that trajectory.

B. Early Indicators Before Visible Overload

InteroSense treats dysregulation as a trajectory rather than a single event. Candidate early indicators may include autonomic deviation from baseline, reduced HRV or delayed recovery after stimulation, rising heart-rate slope during increasing task demand, increased head-motion entropy or repeated orientation shifts, coupling between acoustic load and regulation change, prolonged response latency after transitions, and repeated micro-recovery failures after prompts.

These indicators are not universal biomarkers and must not be interpreted as diagnostic labels. Their meaning depends on child-specific baselines and context. The same heart-rate increase may reflect excitement, frustration, physical movement, anxiety, or recovery depending on the task, therapist annotation, and interaction history. The system therefore estimates trajectories with confidence and explanation rather than issuing hard labels.

C. Personalized Sensory-State Profile and Digital Twin

The personalized sensory-state profile represents how an individual child tends to respond to combinations of sound, movement, proximity, prompt timing, task difficulty, time of day, location, and session structure. The profile is intentionally individualized: one child may show early load during fast auditory prompts, while another may remain stable during

TABLE I
PREDICTION AND ADAPTATION DESIGN MATRIX FOR INTEROSENSE

System layer	Input or evidence stream	Interpretable output	Therapist-supervised adaptation
Earable sensory gateway	PPG-derived heart-rate dynamics, HRV trends, temperature change, head motion, orientation shifts, and acoustic-load descriptors.	Baseline-normalized physiological, movement, and auditory-context features.	Supports early recognition of sensory-state changes without storing raw speech by default.
Robot and session context	Robot speech rate, prompt density, task difficulty, transition timing, movement amplitude, proximity, and therapist annotations.	Explanation of which interaction conditions may be contributing to rising load or recovery difficulty.	Helps adjust pacing, prompt style, sensory channel, transition structure, and recovery windows.
Early-indicator modeling	Short temporal windows combining earable signals with robot/session context.	Confidence-based trajectories such as regulated engagement, rising sensory load, overload risk, disengagement risk, or recovery.	Provides earlier decision support before visible refusal, shutdown, or distress appears.
Personalized sensory-state profile	Child-specific baseline patterns across sessions, environments, tasks, and time of day.	Individualized response patterns rather than universal thresholds.	Enables more personalized session planning and caregiver reporting.
Adaptive sensory digital twin	Updated child-specific state estimates, confirmed or corrected by therapist feedback.	Evolving representation of selected sensory-response patterns relevant to robot interaction.	Supports conservative, explainable, and bounded adaptation over time.

auditory tasks but become dysregulated during rapid transitions or high-motion interaction.

The adaptive sensory digital twin is a dynamic, child-specific computational representation of selected sensory-state patterns relevant to robot interaction. It does not model the whole child and does not claim to know internal experience. It outputs confidence-based estimates and explanations. Early versions should begin with transparent baseline deviation rules and conservative temporal models; later versions can incorporate supervised or semi-supervised learning when enough ethically collected annotations exist. This incremental design keeps the framework scientifically testable and safer for child-facing research.

D. Prediction-to-Adaptation Policy

The adaptation layer is broader than simple robot output control. It can adjust robot micro-behavior, task strategy, session planning, therapist decision support, and caregiver reporting. Examples include reducing speech rate, lowering prompt density, pausing after transitions, switching from auditory prompts to visual choices, inserting predictable recovery windows, changing activity order, or explaining to the therapist which features drove the risk estimate. The robot should not make autonomous high-stakes decisions; it should provide interpretable recommendations that the therapist can accept, edit, or reject.

E. Minimum WIP Feasibility Evidence

InteroSense is currently positioned as a work-in-progress framework. Therefore, the first validation objective is not to claim clinical effectiveness, but to de-risk the sensing, synchronization, comfort, and adaptation pathway before child-facing use. The initial feasibility layer will verify whether the proposed earable form factor can be safely tolerated in short supervised sessions, whether candidate PPG, IMU, temperature, and acoustic-load features can be collected with acceptable signal quality during robot-like interaction tasks, and whether these streams can be synchronized with robot behavior logs and therapist annotations. This minimum validation step is intended to convert the framework from a conceptual architecture into a testable research prototype while preserving ethical boundaries.

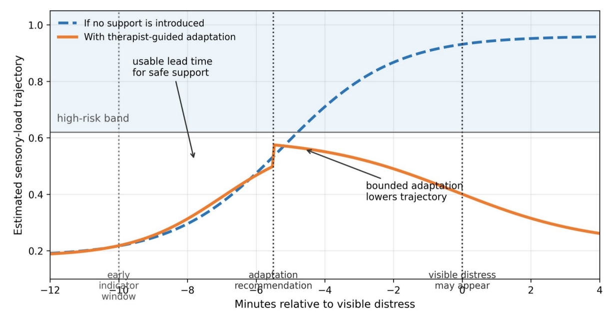

Fig. 3. Conceptual trajectory only, not empirical results. The research question is whether fused earable-plus-context features can provide enough lead time for low-risk, therapist-guided support before visible escalation.

IV. MODELING AND EVALUATION DIRECTION

A practical first implementation should use transparent temporal modeling rather than over-complex clinical claims. Sensor and context streams can be segmented into overlapping windows, for example 30-120 s, then converted into baseline-normalized features. Instead of a binary alarm, the model estimates a state trajectory with confidence and an explanation of the strongest contributing features. Figure 3 shows the target interaction logic: the value of prediction is measured by usable lead time for safe support, not by labeling a child.

The key modeling question is whether fused earable-plus-context features provide earlier, more interpretable support than either modality alone. This requires ablation testing, calibration, false-alarm analysis, and therapist-rated usefulness. Table III gives a staged evaluation plan that avoids clinical overreach.

V. ILLUSTRATIVE INTERACTION SCENARIO

A child begins a robot-assisted sensory-learning activity in a regulated state. The robot introduces a visual categorization task with slow pacing and predictable turn-taking. As the session shifts to faster auditory prompts, the child has not yet shown visible distress. The earable-plus-context model detects a child-specific trajectory: increased heart-rate slope, reduced short-window HRV, repeated head-orientation shifts away from the robot, and slower recovery after prompts while acoustic load and prompt density increase.

TABLE II
MINIMUM WIP FEASIBILITY CHECKS BEFORE CHILD-FACING TESTING

Feasibility concern	**What must be checked**	**Evidence expected at WIP stage**
Earable comfort and fit	Can the device stay stable during short seated interaction?	Adult/bench fit test, CAD render, comfort notes.
Safe removal and hygiene	Can the child or therapist remove it easily? Can it be cleaned?	Design checklist and therapist feedback.
PPG signal quality	Does motion disturb heart-rate/HRV features too much?	Short adult test with rest, speech, and head movement.
IMU/head-motion usability	Can orientation shifts and movement variability be captured?	Basic IMU logging during mock robot tasks.
Acoustic-load privacy	Can sound intensity/variability be captured without storing speech?	Demonstration of coarse audio features only.
Robot-session synchronization	Can sensor windows align with robot prompts, transitions, and task events?	Timestamped sensor and robot-log example.
Battery/heat safety	Does the device remain safe for short supervised sessions?	Bench check before any child-facing session.
Therapist usefulness	Are explanations and recommendations understandable?	Expert review from therapist/HRI researcher.
False-alarm risk	Could the system over-warn or misinterpret excitement as distress?	Planned calibration and therapist override rules.

TABLE III
STAGED EVALUATION PLAN BEFORE CLINICAL CLAIMS

Phase	**Primary question**	**Candidate measures**	**Decision gate**
1. Bench and adult testing	Can the earable prototype be tolerated and produce usable signals during safe non-clinical tasks?	Comfort, fit stability, heat, battery, hygiene, safe removal, PPG quality, motion artifacts.	Prototype is safe enough for expert review and short supervised sessions.
2. Expert review	Do occupational therapists, special educators, caregivers, and HRI researchers find the indicators and actions meaningful?	Action usefulness, explanation clarity, appropriateness of caregiver report, risk concerns.	Adaptation policy is useful and ethically bounded.
3. Child-facing feasibility	After formal ethics approval, can the protocol work in supervised sessions?	Tolerance, signal usability, missed events, false-alarm burden, therapist override frequency.	Feasible before any outcome-efficacy study.
4. Predictive assessment	Do fused features improve earlier and more interpretable support?	AUROC, calibration, lead time, ablations, therapist alignment, generalization limits.	Fused model must outperform earable-only and context-only baselines.

The system does not label the child as distressed. It estimates a rising sensory-load trajectory with moderate confidence and shows the therapist the contributing factors: auditory load, prompt density, and delayed recovery. The recommendation may be to switch from auditory prompts to visual choice-based interaction, reduce prompt density, add a transition preview, insert a predictable recovery pause, and log the pattern for future session design. The therapist remains responsible for interpretation and action.

After the session, the child-specific profile is updated only if therapist feedback supports the interpretation. The caregiver report may state that low-motion visual tasks were tolerated, while repeated fast auditory prompts were associated with early overload indicators and slower recovery. Across sessions, this supports better scheduling, more personalized robot behavior, and clearer communication among therapists and caregivers.

VI. ETHICAL AND PRACTICAL GUARDRAILS

InteroSense involves children, disability-related contexts, biosignals, and adaptive robotics. Predictions must therefore be presented as decision support, not facts about a child's internal state. Confidence, uncertainty, and explanation should remain visible. Data minimization is central: auditory exposure can be represented through coarse acoustic-load features without storing raw speech by default. Parental consent, child assent when appropriate, de-identification, access control, withdrawal procedures, and limits on secondary use are required.

The child must be able to reject or remove the earable, and alternative sensing modes should remain available. The robot should never force interaction, extend a session against the child's needs, or hide its reasoning from the therapist. The appropriate design principle is adaptive but bounded autonomy: the system can support timing, pacing, and strategy selection, while a human professional remains responsible for care decisions.

VII. CONCLUSION

InteroSense positions autism-support robotics within the full human state-aware pipeline: multimodal sensing, temporal state estimation, interpretable modeling, and bounded adaptation. Its distinctive value is the earable-to-robot pathway: physiological, movement, auditory-context, and interaction-temporal streams are fused into child-specific early indicators, updated through an adaptive sensory digital twin, and translated into therapist-guided support across robot behavior, task strategy, session planning, and caregiver reporting.

Future work will focus on prototype development, expert validation, ethics approval, feasibility testing, and careful comparison of context-only, earable-only, and fused predictive models. The long-term direction is a proactive robot-assisted support system that learns how to interact more safely, personally, and transparently with each child over time.

Broader human-centered evaluation should therefore include not only lead time, calibration, and false-alarm burden, but also therapist trust, explanation usefulness, caregiver acceptance, child tolerance, workload reduction, and therapist-robot-caregiver coordination [11]–[14]. Together, these criteria shape a proactive earable-to-robot support system that learns child-specific interaction patterns across repeated sessions while preserving interpretability, calibrated trust, therapist authority, family acceptability, and cultural appropriateness [2], [11], [13].

References

[1] A. Kouroupa, R. Laws, K. Irvine, S. E. Mengoni, A. Baird, and S. Sharma, “The use of social robots with children and young people on the autism spectrum: A systematic review and meta-analysis,” *PLOS ONE*, vol. 17, no. 6, e0269800, 2022.

[2] B. Scassellati *et al.*, “Improving social skills in children with ASD using a long-term, in-home social robot,” *Science Robotics*, vol. 3, no. 21, eaat7544, 2018.

[3] H. Javed, R. Burns, M. Jeon, A. M. Howard, and C. H. Park, “A robotic framework to facilitate sensory experiences for children with autism spectrum disorder: A preliminary study,” *ACM Transactions on Human-Robot Interaction*, vol. 9, no. 1, pp. 1–26, 2019.

[4] A. Q. Alban *et al.*, “Heart rate as a predictor of challenging behaviours among children with autism from wearable sensors in social robot interactions,” *Robotics*, vol. 12, no. 2, Art. 55, 2023.

[5] Y. Kartha *et al.*, “Prediction of challenging behaviors associated with profound autism in a classroom setting using wearable sensors,” arXiv:2605.17618, 2026.

[6] M. Zwilling, A. Romano, H. Hoffman, M. Lotan, and R. Tesler, "Development and validation of a system for the prediction of challenging behaviors of people with autism spectrum disorder based on a smart wearable shirt: A mixed-methods design," *Frontiers in Behavioral Neuroscience*, vol. 16, Art. 948184, 2022.

[7] V. Goverdovsky *et al.*, “Hearables: Multimodal physiological in-ear sensing,” *Scientific Reports*, vol. 7, Art. 6948, 2017.

[8] H. Barki and W.-Y. Chung, “Mental stress detection using a wearable in-ear plethysmography,” *Biosensors*, vol. 13, no. 3, Art. 397, 2023.

[9] C. Hu *et al.*, “A survey of earable technology: Trends, tools, and the road ahead,” arXiv:2506.05720, 2025.

[10] A. Fuller, Z. Fan, C. Day, and C. Barlow, “Digital twin: Enabling technologies, challenges and open research,” *IEEE Access*, vol. 8, pp. 108952–108971, 2020.

[11] E. J. de Visser *et al.*, “Towards a theory of longitudinal trust calibration in human-robot teams,” *International Journal of Social Robotics*, vol. 12, no. 2, pp. 459–478, 2020.

[12] L. Sanneman and J. A. Shah, “Trust considerations for explainable robots: A human factors perspective,” arXiv:2005.05940, 2020.

[13] V. Lim, M. Rooksby, and E. S. Cross, “Social robots on a global stage: Establishing a role for culture during human-robot interaction,” *International Journal of Social Robotics*, vol. 13, no. 6, pp. 1307–1333, 2021.

[14] S. Habayeb *et al.*, “Meeting the needs of autistic adults in Qatar: Stakeholder perspectives on gaps in services and priorities for future programming,” *Autism*, vol. 26, no. 1, pp. 88–100, 2022.