Beyond Fixed Loops: Estimated RPE as a Master Signal for Timely Proactive Coaching in Physical Workouts.

Thomas Baldassarre[1,*,†], Evelyne Meurisse[1], Kevin El Haddad[1] and Thierry Dutoit[1]

[1] *Numediart Institute, ISIA Lab, University of Mons, Mons, Belgium*

Abstract

While proactive Generative AI (GenAI) assistants promise to transform productivity, they are currently bottlenecked by fixed interaction loops that ignore the user's internal state, leading to a 'Hyper-Proactivity Paradox' where poorly timed interventions disrupt cognitive or physical flow. A critical driver of this failure is the 'Interaction Tax.' Research by Irfan et al.[1] and Zafrani et al. [2] established that manual exercise data logging, particularly via tablets, is frequently perceived as a barrier to engagement, highlighting the need for passive sensing to maintain user adherence. This paper proposes estimated RPE eRPE as a latent state estimation problem to resolve this timing gap. By fusing multi-modal data—including biomechanical velocity decay (v_{mean}), cardiovascular drift, and affective valence—we introduce a pipeline that enables agents to transition from reactive monitors to intelligent, context-aware partners. Our framework facilitates 'Contextual Silence,' allowing the agent to preserve social capital and only deliver high-density affective scaffolding during the physiological or cognitive inflection point.

Keywords

Socially Assistive Agents, Estimated RPE, Proactive AI, Interaction Tax, Contextual Silence

1. Introduction

The evolution of Generative AI (GenAI) from reactive responders to proactive partners has hit a fundamental bottleneck: the Timing Gap. While LLMs have begun to anticipate semantic intent in knowledge work, Socially Assistive Agents (SAA) in high-stakes, embodied contexts—such as sports and physical rehabilitation—remain tethered to Fixed Temporal Loops. This creates a systemic ergonomic failure where agents deliver feedback at rigid intervals, ignoring the non-linear nature of human fatigue. The result is a Hyper-Proactivity Paradox: interventions delivered independently of physiological context are perceived as intrusive nuisances, directly undermining the agent's perceived competence and the user's trust.

As established by Irfan et al.[1], manual data capture tools—such as tablet-based logging—are frequently perceived by users as a significant burden compared to automated systems, leading to lower adherence even in essential health and rehabilitation contexts. In an embodied performance state, this burden translates into a prohibitive 'Interaction Tax' that destroys the physiological flow necessary for peak performance. Traditional tools, such as the Borg Rating of Perceived Exertion (RPE), require users to pause their physical set to verbally report their internal state. This forces a jarred transition from an embodied, kinesthetic state to a declarative interface state, effectively destroying the "Flow" necessary for peak performance. Longitudinal evidence (2020-2026) suggests that this tax is exacerbated by system latency; response delays exceeding 4 to 5 seconds during manual reporting trigger acute user frustration.

To resolve this, we propose a paradigm shift toward Computational Interoception through the concept of estimated RPE eRPE. Rather than treating exertion as a subjective report, we frame it as a Latent State

Proceedings of TimelyAI: When Should Generative AI Assistants Intervene? A Workshop Co-located with the 5th Symposium on Human-Computer Interaction for Work (CHIWORK 2026), Linz, Austria, June 22 to 25, 2026

* Corresponding author.
† These authors contributed equally.
thomas.baldassarre@umons.ac.be (T. Baldassarre); evelyne.meurisse.ac.be (E. Meurisse)

Estimation problem. By fusing three objective data streams—biomechanical velocity decay (v_{mean}), cardiovascular drift, and affective valence—we can infer high-level psychophysiological strain in real-time. This allows for the implementation of Contextual Silence: a policy where the agent preserves its "social capital" during low-exertion phases and only provides high-density affective or technical scaffolding.

In this paper, we provide a technical blueprint for exertion-aware agents that gate interventions based on functional necessity rather than temporal thresholds. By eliminating the Interaction Tax through a VBT-RIR-RPE pipeline, we demonstrate how proactive AI can maintain a high-integrity therapeutic alliance without inducing cognitive or kinesthetic friction.

2. Literature Review

A meta-analysis of longitudinal data (2020–2026) demonstrates that while SAA achieve superior physical results compared to traditional interfaces, their underlying "timing logic" remains primitive. Current proactivity methodologies rely on three flawed socio-technical pillars that impede the establishment of a truly adaptive athletic or therapeutic alliance.

2.1. The Failure of Static Temporal Interaction Loops

The dominant paradigm in current SAA research is the Temporal and Cyclic Feedback Model. In this framework, the moment of intervention is treated as a fixed technical parameter rather than a dynamic variable.

Céspedes et al. [3] demonstrate that robot-led cardiac rehabilitation increases adherence by 13.3% and improves recovery heart rate, with patients maintaining engagement over 18 weeks. However, as noted by Zafrani et al.[2], a subset of 'Skeptical' users still risks perceiving the agent as 'nagging' if the intervention timing remains tied to fixed loops rather than real-time struggle

2.2. The Resolution Gap in Reactive Physiological Triggers

Advanced systems attempt to solve the timing problem by integrating real-time biological signals, yet these remain predominantly Safety-Oriented rather than Support-Oriented.

- Binary Thresholds: Existing systems utilize heart rate (HR) sensors to trigger instructions only when a user exceeds a safe training limit, typically 85% of their theoretical maximum HR.
- Lack of Spontaneous Responsivity: While effective for risk mitigation, these triggers lack the signal resolution required to distinguish between high-intensity peak performance ("Flow") and acute psychological disengagement. As noted by Platz et al.[4], current humanoid agents lack "responsivity to spontaneous cues," leading them to either remain contextually silent during a user's performance inflection point. or interrupt a productive set with technical corrections

2.3. The Prohibitive "Interaction Tax" of Manual Reporting

To mitigate the lack of physiological signal resolution, many researchers employ a Declarative and Manual Reporting Model, such as the verbal solicitation of the Borg Rating of Perceived Exertion.

- Flow Disruption: This methodology imposes a severe "Interaction Tax". In robot-assisted gait rehabilitation, pausing the session to solicit user input forces a jarred transition from a kinesthetic, embodied state to a declarative, interface-oriented state.
- The Usability Paradox: Even in systems with positive usability ratings—such as the 77.5 SUS score reported by Haeussl et al. [5] for the robot Pepper—fixed feedback frequencies can trigger frustration. This 'Interaction Tax' is exacerbated by system latency; Feingold-Polak and Levy-Tzedek [6] established that response delays exceeding 4 to 5 seconds during manual reporting cause acute user frustration and disrupt 'Flow'.

3. The Proposed Paradigm: eRPE as a Latent State Estimation

The central thesis of this position paper is that the "Timing Gap" in proactive coaching can only be resolved by transitioning from declarative reporting to a model of Computational Interoception. Current SAA are "blind" to the user's internal state, lacking the integrated sensing required to detect muscle innervation or fatigue-induced velocity loss. We propose the concept of eRPE as a Latent State Estimation problem—the challenge of inferring high-level psychophysiological strain through the fusion of multi-modal kinematic and cardiovascular signals.

3.1. The Ergonomic Imperative: Eliminating the Interaction Tax

In Resistance Training (RT), pausing a set to solicit a subjective Borg Scale rating imposes a prohibitive "Interaction Tax". The estimation of RPE is a mandatory ergonomic evolution; it allows the agent to establish Contextual Intelligence by observing performance passively, thereby maintaining the user's embodied state

3.2. Logical Architecture: The VBT-RIR-RPE Pipeline

To bridge the resolution gap, we hypothesize that eRPE must be estimated through the concurrent analysis of three objective data streams (Figure 1), with a primary focus on the inverse relationship between movement velocity and exertion:

- Biomechanical Velocity Decay (VBT): The eRPE model utilizes kinematic signatures—specifically velocity decay derived from Velocity-Based Training (VBT) metrics—as a proxy for affective valence. While VBT traditionally monitors physical capacity, foundational work by Giraud et al. [7] demonstrates that high-level movement qualities such as 'energy' and 'impulsivity' are reliable predictors of emotional dimensions. By mapping these kinematic fluctuations to internal states, the system can leverage findings from Haeussl et al. [5] regarding the ability of contemporary AI to analyze body posture and voice tone. This enables the agent to provide timely, empathic coaching interventions even when facial signals are obscured by the physical intensity of the workout, effectively bypassing the 'interaction tax' of manual reporting.
- Cardiovascular Drift: To quantify global physiological cost, the eRPE framework integrates Heart Rate Variability (HRV) and cardiovascular drift into its predictive pipeline. While VBT identifies local neuromuscular fatigue, HRV provides the systemic "biological anchor" required to distinguish between high-intensity flow and acute over-exertion.
- Affective Valence: The eRPE model incorporates facial emotion recognition (FER) to distinguish between 'productive struggle' and acute distress. This is validated by Naseer et al. [8], who demonstrated that hybrid AI architectures—combining transformer models (BERT) for linguistic context with CNN-LSTMs for facial expression analysis —can successfully infer a user's emotional state by tracking temporal changes during assistive tasks. This supports the eRPE pipeline's ability to detect the transition from 'healthy effort' to 'over-exertion' without requiring manual interruption , thereby bypassing the 'interaction tax' of self-reporting.

3.3. Theoretical Research Agenda: Calibration and Constraints

In defining eRPE as the "missing link" for Timely AI, we identify three critical research challenges:

- The Interoceptive Calibration Gap: How to calibrate algorithms to recognize that a novice's velocity decay might signal psychological quitting, while an expert's v_{mean} crash represents a true physiological limit.
- Boundary Constraints in Sensing: While VBT is highly effective for linear resistance movements, research must address the sensing of eRPE in isometric or non-linear movements where velocity is not a viable master signal

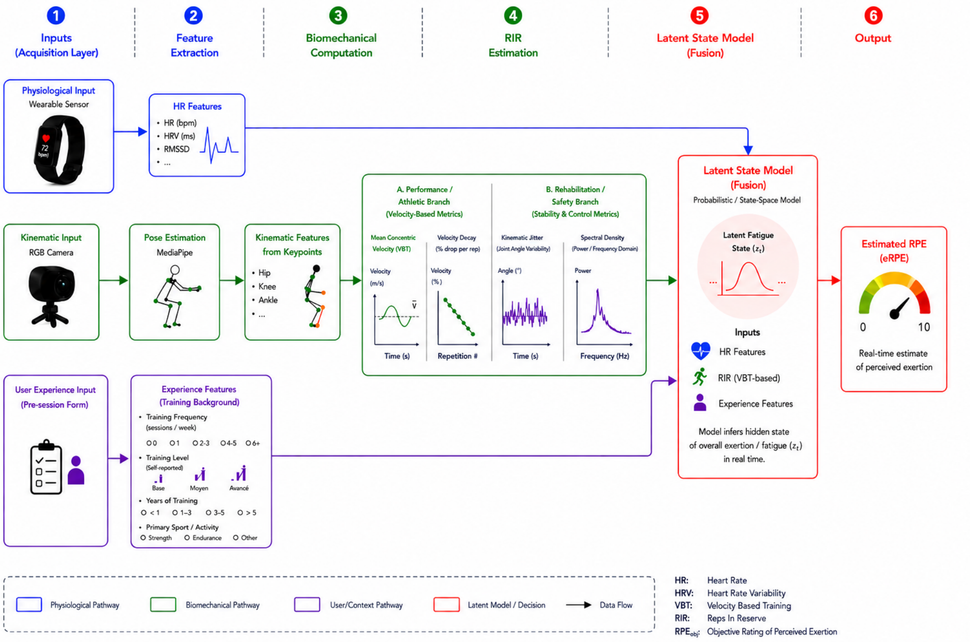

Figure 1: Multimodal pipeline for real-time RPE estimation (eRPE) combining physiological, kinematic, and contextual features through a latent state model

- The Multimodal Proactive Orchestration Pipeline: Investigating the integration of eRPE into a multi-layered decision pipeline that governs agent behavior. This research must define a fusion architecture where biomechanical (VBT), physiological (HRV), and affective signals are synchronized to determine the agent's "Social Presence".

4. FUNCTIONAL PROPOSITION: THE EFFORT-SEMANTIC-TIMING CORRELATION

We propose a multi-layered decision-making model that synchronizes the agent's intervention frequency and semantic complexity with the user's real-time effort curve. This framework, grounded in the eRPE signal, facilitates a transition from "instructional dominance" to Contextual Silence, directly addressing the "repetitiveness" and "social skill" failures identified in the longitudinal cardiac studies of Céspedes et al. [3] and Irfan et al.[1]

4.1. Phase A: Validation and the Preservation of Social Capital (Low eRPE)

During the initiation of an exercise set (high Reps In Reserve (RIR)), the agent's primary strategy is Contextual Silence. By implementing a policy of Communicative Economy during high-RIR phases, the agent avoids Cognitive-Kinesthetic Friction and preserves its 'social capital' for the terminal repetitions of the set. In our proposed model, the agent monitors kinematic accuracy via skeletal tracking; if form matches the prescribed template, the agent remains silent or provides discrete, positive validation (e.g., "Perfect posture") at a minimal frequency. This strategy preserves the agent's "social capital," ensuring that when it finally speaks, the intervention carries the "authority" required for behavioral change.

4.2. Phase B: Technical Scaffolding and Competence Building (Moderate eRPE)

As the eRPE identifies initial physiological strain—signaled by a rise in heart rate variability and subtle kinematic jitter—the agent transitions to Low-Frequency, High-Specificity technical feedback. This phase is critical for novices who lack Interoceptive Calibration and are prone to early abandonment or injury. Rather than generic praise, which Rea et al.[9] identified as potentially "insincere" and uncompelling, the agent provides "rich information" grounded in the user's current performance (e.g., "10 reps done, maintain back alignment"). By establishing the agent as a competent technical partner, we foster the Therapeutic Alliance necessary for long-term adherence.

4.3. Phase C: Affective Scaffolding (High eRPE)

The terminal phase represents the peak of the effort curve where RIR reaches zero and the user enters the "biological burn". At this inflection point, the agent's decision-logic must diverge based on the Experience Profile ($P_{profile}$) to ensure either performance optimization or clinical safety.

4.3.1. Hypertrophy and Athletics: Performance Scaffolding

To prevent premature abandonment, the agent transitions to High-Frequency, Low-Complexity Affective Coaching. As established by Vankit et al.[10], the combination of multimodal strategies (subtitles and visualizations) in long VR conversations significantly increases mental and temporal workload (NASA-TLX). Therefore, during peak eRPE, the agent must reduce Linguistic Density to avoid overwhelming the user's remaining cognitive resources.

- Trigger: A Mean Concentric Velocity (v_{mean}) crash exceeding 30% of the set's initial pace.
- Behavior: The agent abandons complex technical cues in favor of rhythmic, low-entropy verbal support (e.g., "Push through," "Final two") to preserve the athletic alliance without inducing cognitive-kinesthetic friction.

4.3.2. Clinical Rehabilitation: The "Invisible Guardrail"

In rehabilitation, where movements are intentionally slow ($v_{planned}$), eRPE serves as a safety-critical monitor rather than a motivational catalyst. The goal is to keep the patient within the Optimized Therapeutic Window—maximizing effort without breaching the threshold of acute pain or injury.

- Trigger (Kinematic Jitter): Instead of simple velocity decay, the agent monitors for high-frequency neuromuscular tremor. By analyzing the spectral density of the 3D skeletal path, the agent identifies micro-oscillations that signal the onset of localized muscle failure or pain
- Trigger (Compensatory Mechanics): The agent detects subtle deviations from the prescribed template, such as trunk leaning or asymmetric limb loading
- Behavior: This automated 'invisible guardrail' utilizes the high-resolution tracking capabilities described in the recent framework by Naseer et al.[8]—specifically the use of Optical Flow Analysis to detect posture shifts and movement instability in real-time. By leveraging this technical anchor, our system can intervene the moment Kinematic Fluency degrades into jitter, ensuring the user remains within an optimized therapeutic window without response latencies.

4.4. Adaptive Representation and Presence Density

Finally, our proposition argues that the Avatar Representation must reflect this temporal logic. Tan et al.[11] and Rings et al.[12] suggest that human-like interactions enhance engagement, yet their "humanness" also raises user expectations for empathy. By linking presence density to effort, the agent avoids the "nagging" profile identified by Zafrani et al.[2]. By linking presence density to effort, the agent becomes a truly Empathetic Partner. As Chauvin et al.[13] demonstrated, non-directive, empathetic virtual agents are perceived as more trustworthy and significantly raise a user's self-efficacy to overcome barriers compared to directive interventions.

5. CONCLUSION: TOWARD "CONTEXTUALLY SILENT" AI

The fundamental objective of this research is to catalyze a paradigm shift in the architecture of SAA by transitioning from subjective, declarative measurement models to an objective, computational sensing framework. Current frameworks, as critically analyzed in the longitudinal work of Céspedes et al.[3], are bottlenecked by the requirement for periodic manual prompts, which impose a prohibitive "Interaction Tax" on the user. By replacing these intrusive self-reports with a continuous invisible stream of eRPE data, agents can move beyond the "Alarm-Clock" model toward a state of functional and social autonomy. Our proposed eRPE framework, currently in the theoretical formulation phase and conceptual reflection phase, seeks to integrate physiological and kinematic markers to adapt agent behavior based on three critical variables: user experience profile ($P_{profile}$), specific therapeutic objectives and real-time exertion.

Although this methodology is currently tailored to high-intensity athletic and rehabilitation contexts, its success serves as a vital proof-of-concept for the broader evolution of agent responsivity. As established by Platz et al.[4], humanoid agents currently do not respond to spontaneous cues, a failure that Zafrani et al. [2] demonstrates leads to a "Novelty Threshold" where users dismiss the agent as a mechanical nuisance once the initial excitement dissipates. By integrating experience profiles to distinguish between novices and professionals, an autonomous agent can finally achieve Contextual Silence—remaining invisible when effort is managed effectively and providing high-intensity affective support only during the biological performance inflection point. This shift directly addresses the "repetition" and "lack of social skill" failures documented across the current SAA landscape.

The development of the "ideal" coaching agent requires moving away from the constant verbal loops of early-generation robotics toward a "Just-in-Time" communicative logic. As the studies of Tan et al.[11] and Rings et al.[12] indicate, users prioritize technical precision and relatable social presence, yet are frequently deterred by the excessive cognitive load of multimodal over-correction. By grounding interventions in the "rich information" provided by the master signal eRPE, every spoken word becomes functionally necessary—whether it is to navigate a beginner through an injury-prone state or push an athlete through a terminal struggle. The success of future proactive AI will not be measured by the quantity of its discourse, but by its ability to remain Contextually Silent until the exact physiological inflection point required to sustain the therapeutic alliance. Future research must now address the Ethical Guardrails for 'invisible' sensing, ensuring that continuous biological monitoring is balanced with user transparency.

Acknowledgments

The authors express their gratitude to the colleagues at UMONS University for their technical insights. This work was supported by funding from the Wal4XR project.

Declaration on Generative AI

During the preparation of this work, the authors used Llama 4 in order to: Grammar and spelling check. Furthermore, the authors utilized nano banana to refine and enhance the visual schema presented in Figure 1. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] B. Irfan, N. Céspedes, J. Casas, E. Senft, L. F. Gutiérrez, M. Rincon-Roncancio, C. A. Cifuentes, T. Belpaeme, and M. Múnera. Personalised socially assistive robot for cardiac rehabilitation:

Critical reflections on long-term interactions in the real world: Personalised robot for cardiac rehabilitation. *User Modeling and User-Adapted Interaction*, 33(2):82698–82717, 2022.

[2] O. Zafrani, G. Nimrod, M. Krakovski, S. Kumar, S. Bar-Haim, and Y. Edan. Assimilation of socially assistive robots by older adults: an interplay of uses, constraints and outcomes. *Frontiers in Robotics and AI*, 11, 2024.

[3] N. Céspedes, B. Irfan, E. Senft, C. A. Cifuentes, L. F. Gutierrez, M. Rincon-Roncancio, T. Belpaeme, and M. Múnera. A Socially Assistive Robot for Long-Term Cardiac Rehabilitation in the Real World. *Frontiers in Neurorobotics*, 15, 2021.

[4] T. Platz, A. L. Pedersen, P. Deutsch, A.-N. Umlauft, and S. Bader. Analysis of the therapeutic interaction provided by a humanoid robot serving stroke survivors as a therapeutic assistant for arm rehabilitation. *Frontiers in Robotics and AI*, 10, 2023.

[5] A. Haeussl, I. Zwigl, L. Stojec, I. Smolak, M. Stijic, T. Stross, M. Lenger, F. T. Fellendorf, S. Guggemos, E. M. D. Schoenthaler, M. Pszeida, T. Orgel, S. Draxler, M. Schneeberger, S. Russegger, J. Zuschnegg, D. Steindl, A. Schultz, S. Schuessler, M. Macher, L. Paletta, N. Dalkner, and E. Z. Reininghaus. Evaluating usability and acceptance of a socially assistive robot supported cognitive training for depression – results of the randomized controlled pilot study 'AMIGA'. *Frontiers in Psychiatry*, 16, 2026.

[6] R. Feingold Polak and S. Levy-Tzedek. Insights from a Long-Term in-the-Wild Study with Post-Stroke Patients using a Socially Assistive Robot. In *Proceedings of the 5th International Conference on Computer-Human Interaction Research and Applications (CHIRA 2021)*, pages 319–323. SCITEPRESS, 2021.

[7] T. Giraud, F. Focone, V. Demulier, J. C. Martin, and B. Isableu. Perception of Emotion and Personality through Full-Body Movement Qualities: A Sport Coach Case Study. *ACM Transactions on Applied Perception*, 13(1), 2015.

[8] F. Naseer, M. N. Khan, M. Tahir, A. Addas, and H. Kashif. Enhancing Elderly Care With Socially Assistive Robots: A Holistic Framework for Mobility, Interaction, and Well-Being. *IEEE Access*, 13:82698–82717, 2025.

[9] D. J. Rea, S. Schneider, and T. Kanda. "Is this all you can do? Harder!": The Effects of (Im)Polite Robot Encouragement on Exercise Effort. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, 2021. ACM.

[10] S. A. Vankit, S. Zamanifard, D. Diaz, C. Mousas, K. Richardson, A. T. Duchowski, and M. Volonte. Exploring the Impact of Multimodal Long Conversations in VR on Attitudes Toward Behavior Change, Memory Retention, and Cognitive Load. *Computer Animation and Virtual Worlds*, 2025.

[11] C. T. Tan, I. Atmosukarto, B. Tandianus, S. Shen, and S. Wong. Exploring the Impact of Avatar Representations in AI Chatbot Tutors on Learning Experiences. In *Proceedings of the 2025 ACM SIGCHI Symposium on Computer-Human Interaction in Play*, New York, NY, USA, 2025. ACM.

[12] S. Rings, L. Kruse, C. Rudschies, L. Rolvien, G. Schauenburg, J. Gallinat, I. Schneider, and F. Steinicke. IVAs' Future in Therapy: Investigating the Impact of Intelligent Virtual Agents in VR on Psychological Self-Help Exercises and User Satisfaction. In *Proceedings of the 2025 International Conference on Intelligent Virtual Agents*, New York, NY, USA, 2025. ACM.

[13] R. Chauvin, C. Clavel, N. Sabouret, and B. Ravenet. A virtual coach with more or less empathy: impact on older adults' engagement to exercise. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, New York, NY, USA, 2023. ACM.