Shall We Dance? Resonance of Intentions with an Embodied Agent based on the Free Energy Principle

TAKERU HASHIMOTO, Sony Computer Science Laboratories, Japan
JUN TANI, Okinawa Institute of Science and Technology, Japan
SHUNICHI KASAHARA, Sony Computer Science Laboratories, Japan

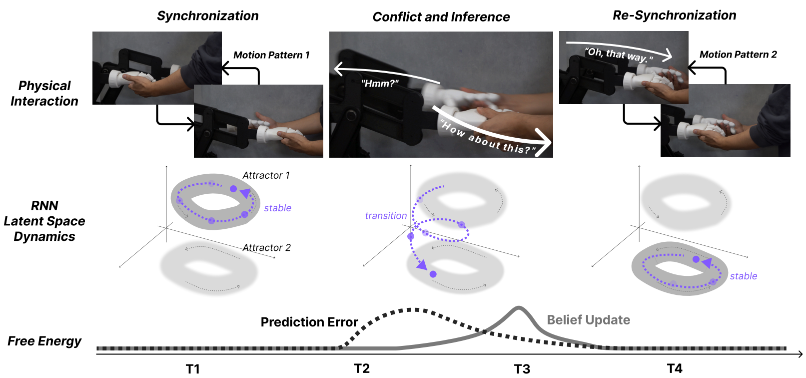

Fig. 1. Real-time adaptive physical coordination via latent attractor dynamics driven by free energy minimization. The robot agent and human perform a continuous hand-in-hand dancing. The process illustrates the agent's adaptation to a human partner's changing intention through three phases: (T1) Synchronization: They share a synchronized rhythm, represented as a stable limit cycle (Attractor 1) in the RNN latent space. (T2) Conflict: When the human changes intention, the physical mismatch generates a spike in Prediction Error in the agent brain (dashed line). (T3) Inference: To minimize this prediction error, the system infers how to change their belief (i.e., Belief Update, solid line). This corresponds to driving the internal state to transition from the current attractor to a new one (Attractor 1 ⟶ 2). (T4) Re-synchronization: The internal state settles into the new attractor, allowing the robot to physically adapt to the human's new intention.

Physical joint action fosters a profound sense of partnership by enabling individuals to negotiate intentions through mutual coordination. However, conventional physical agents typically rely on pre-learned transition rules, limiting their ability to adapt to a partner's improvisational interventions. To address this, we propose a physical agent grounded in the Free Energy Principle that negotiates intentions through physical interaction. Utilizing a predictive coding-based RNN, the agent dynamically updates its internal intention (μ) and time constant (τ) to minimize real-time prediction errors caused by physical contact. This mechanism allows the agent to synchronize with the human's choreography and tempo. In the demonstration "Shall We Dance?", visitors experience this synchronization directly through physical contact with an agent. When intentions align, the initial resistance at the hands is eliminated, providing a tactile confirmation that the agent has successfully inferred the user's new intent without any verbal communication.

Authors' Contact Information: Takeru Hashimoto, hashimoto@csl.sony.co.jp, Sony Computer Science Laboratories, Tokyo, Japan; Jun Tani, jun.tani@oist.jp, Okinawa Institute of Science and Technology, Okinawa, Japan; Shunichi Kasahara, kasahara@csl.sony.co.jp, Sony Computer Science Laboratories, Tokyo, Japan.

Manuscript submitted to ACM

Additional Key Words and Phrases: Physical Human-Robot Interaction, Joint Action, Free Energy Principle, Predictive Coding, Active Inference

ACM Reference Format:
Takeru Hashimoto, Jun Tani, and Shunichi Kasahara. 2026. Shall We Dance? Resonance of Intentions with an Embodied Agent based on the Free Energy Principle. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Human joint action involves continuous negotiation, where partners maintain their own intentions (priors) while inferring those of others [3]. This coordination can enhance motor performance beyond individual capabilities in terms of precision and efficiency. However, most conventional physical agents treat interaction errors merely as control problems to be resolved through motor output. Consequently, these systems lack the cognitive capacity to lead with intent or yield with understanding.

In cognitive robotics, predictive coding-inspired Variational RNNs have been proposed to realize such cognitive give-and-take based on free energy minimization [1]. While these models enable robots to mimic behaviors like turn-taking [5] and mutual yielding [4], they typically rely on transitions learned as statistical regularities from training data. Consequently, these agents struggle to adapt to improvisational interventions that occur at arbitrary moments.

To address this, we propose a system grounded in the Free Energy Principle (FEP) that eliminates the need for learned switching rules. Instead, intention switching emerges dynamically through real-time prediction error minimization during physical contact. This allows the agent to follow a partner's changes from any state, regardless of prior training. We demonstrate this "resonance" using a hand-in-hand dance setup between a human and a custom-made physical agent.

2 System Architecture

2.1 Predictive Coding-based RNN Model

The core of our system is a continuous-time RNN (CTRNN) based on the concept of predictive coding [2]. To enable the agent to autonomously generate multiple movement patterns as prior beliefs, we embed multiple limit cycles as attractors in the high-dimensional latent space through training (Fig. 2 top).The internal state $\boldsymbol{\mu}_t \in \mathbb{R}^N$ ($N = 64$) represents the expected state of the hidden dynamics, updated as:

$$\begin{aligned} \boldsymbol{\mu}_t &= \left(1 - \frac{1}{\tau}\right)\boldsymbol{\mu}_{t-1} + \frac{1}{\tau}(\mathbf{W}_h\mathbf{h}_{t-1} + \mathbf{b}_h) \\ \mathbf{h}_t &= \tanh(\boldsymbol{\mu}_t) \end{aligned} \tag{1}$$

where τ is the time constant, $\mathbf{W}_h$ is the recurrent weight matrix, $\mathbf{b}_h$ is the bias vector, and $\mathbf{h}_t$ is the activation value. Under the Free Energy Principle, we treat the observed movement $\mathbf{x}_t$ and state $\boldsymbol{\mu}_t$ as Gaussian random variables. During training, the network minimizes the Variational Free Energy $\mathcal{F}$, approximated by the sum of squared prediction errors:

$$\mathcal{F} \approx \underbrace{\frac{\pi_x}{2}|\mathbf{x}_t - g(\boldsymbol{\mu}_t)|^2}_{\text{Prediction Error } E_x} + \underbrace{\frac{\pi_\mu}{2}|\boldsymbol{\mu}_t - f(\boldsymbol{\mu}_{t-1})|^2}_{\text{State Error } E_\mu} \tag{2}$$

where $g(\cdot)$ represents the decoder mapping latent states to physical coordinates, and $f(\cdot)$ denotes the transition dynamics defined in Eq. (1).

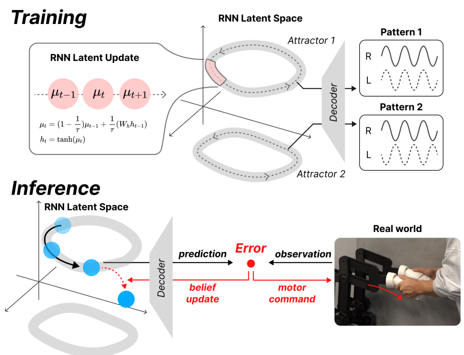

Fig. 2. Schematic of the proposed RNN-based active inference model. **Training (top):** The network learns multiple motion primitives as distinct attractors in the latent space. **Inference (bottom):** When a movement gap occurs, the resulting prediction error triggers a belief update. By shifting between attractors to minimize this error, the agent achieves real-time synchronization.

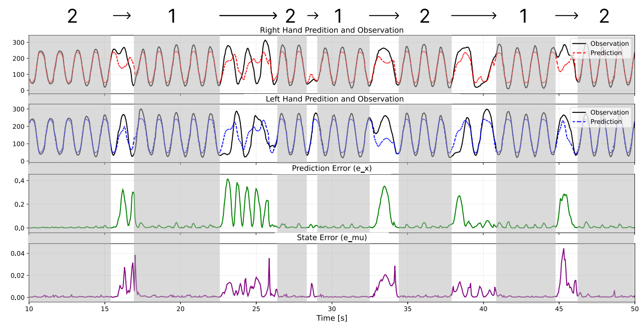

Fig. 3. Time-series of motion switching via physical interaction. The top two plots show the trajectories of the right and left hands (solid line: observation, dashed line: agent's prediction). The bottom two plots show the Prediction Error (E_x) and State Error (E_μ). Gray shaded regions indicate the periods where the human partner applied force to switch the patterns (e.g., from Pattern 2 to Pattern 1).

2.2 Inference: Real-time Intention and Tempo Negotiation

In the inference phase, the model weights are fixed, and the system optimizes $\boldsymbol{\mu}$ and τ in real-time against sequential observations $\mathbf{x}$. By minimizing $\mathcal{F}$, the agent calculates the optimal $\boldsymbol{\mu}$ that explains the partner's movement while maintaining its current prior. This allows the agent to adjust its phase or switch between attractors by crossing the "saddle" when the partner's movement deviates from the current prediction. Simultaneously, by updating the time constant τ, the agent adapts its execution speed to match the partner's tempo.

2.3 Action-Perception Loop via Custom Robot Arm

The interaction is realized through the simultaneous processes of Action and Perception (Fig. 2 bottom). In Perception (belief update), the agent shifts its latent state $\boldsymbol{\mu}_t$ to infer the partner's intention. In Action, the agent generates motor commands via a custom-made 2-DOF force-feedback robot arm. This in-situ loop enables the agent to transition from "leading" to "yielding" based on real-time prediction errors. The physical result is a "resonance of intentions," where resistance vanishes as the system synchronizes with the human partner.

3 Technical Validation

We validated the active inference capability by testing whether the agent could update its intention via external intervention. While the agent initially generated random motions, a human partner applied force to guide it toward target trajectories, alternating between Pattern 1 and Pattern 2. Fig. 3 illustrates the adaptive process via physical interaction. When a human applies force (white areas), the Prediction Error E_x initially rises due to the sensory-motor discrepancy. To minimize this, the agent updates its internal state, causing a transient increase in State Error E_μ. As the agent adapts to the new pattern, both errors decrease, and the system re-synchronizes (gray areas). This sequence—from external force to state update and re-settling—confirms that the inference mechanism effectively adapts the agent's intention in real-time.

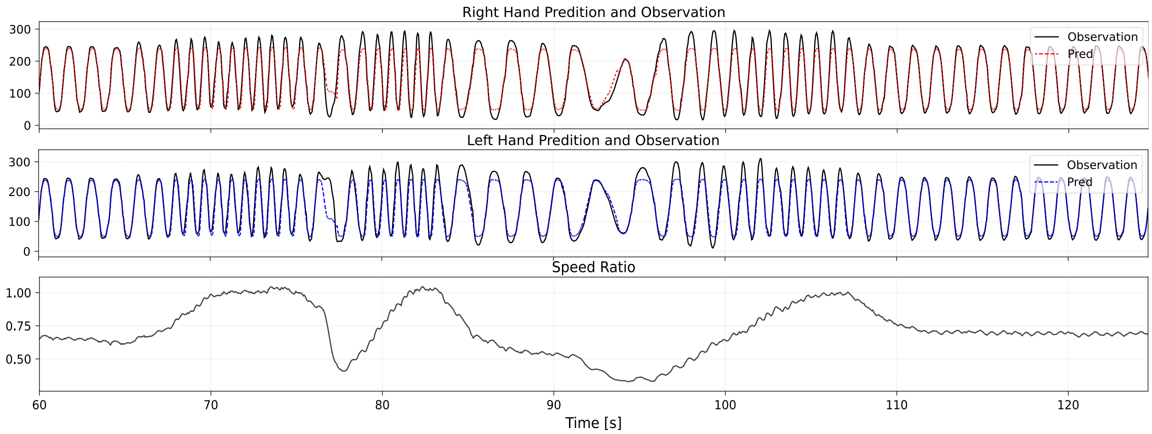

Fig. 4. Time-series log of speed adaptation. The top two plots show the trajectories of the Right and Left hands (solid line: observation, dashed line: prediction). The bottom plot shows the transition of the inferred Speed Ratio (α). As the partner changes the interaction frequency, the agent updates α to synchronize the shared tempo.

To achieve temporal coordination (e.g., "groove"), we treat motion speed as an inferable variable by introducing a speed ratio α that modulates the time constant τ. By updating the effective time constant as τ/α, the agent can accelerate or decelerate its internal dynamics to match the partner. Fig. 4 demonstrates this adaptation as the partner changes the interaction frequency. The agent infers the optimal α to minimize temporal prediction errors. As shown in the bottom plot, α dynamically tracks the partner's speed, allowing the agent to synchronize its tempo while maintaining the original attractor structure. This confirms that the model can flexibly adapt to varying rhythms through real-time inference of the time constant.

4 Demonstration: Shall we Dance?

In this installation, experimenter hold hands with the robot to dance. The core of the experience is the process of two intentions merging into one. When synchronized, you feel a "shared breath" where movements flow together perfectly. However, if you change the choreography, a gap between intentions creates a brief resistance. As the agent understands your new intent, this friction vanishes into lightness. By feeling this “resonance of intentions”, the shift from conflict to harmony, you realize your heart has been understood without a single word.

References

[1] Ahmadreza Ahmadi and Jun Tani. 2019. A novel predictive-coding-inspired variational RNN model for online prediction and recognition. *Neural Comput.* 31, 11 (Nov. 2019), 2025–2074. doi:10.1162/neco_a_01228

[2] Karl Friston. 2005. A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1456 (April 2005), 815–836. doi:10.1098/rstb.2005.1622

[3] G Ganesh, A Takagi, R Osu, T Yoshioka, M Kawato, and E Burdet. 2014. Two is better than one: physical interactions improve motor performance in humans. *Sci. Rep.* 4 (Jan. 2014), 3824. doi:10.1038/srep03824

[4] Hiroki Sawada, Wataru Ohata, and Jun Tani. 2025. Human-robot kinaesthetic interactions based on the free-energy principle. *IEEE Trans. Syst. Man Cybern. Syst.* 55, 1 (Jan. 2025), 366–379. doi:10.1109/tsmc.2024.3481251

[5] Nadine Wirkuttis, Wataru Ohata, and Jun Tani. 2023. Turn-taking mechanisms in imitative interaction: Robotic social interaction based on the free energy principle. *Entropy (Basel)* 25, 2 (Jan. 2023), 263. doi:10.3390/e25020263