Check for updates

# Neuroadaptive haptics: a proof-of-concept comparing reinforcement learning from explicit ratings and neural signals for adaptive XR systems

Lukas Gehrke[1]*, Aleksandrs Koselevs[1], Marius Klug[2] and Klaus Gramann[1]

[1]Department of Biopsychology and Neuroergonomics, Technische Universität Berlin, Berlin, Germany,
[2]Young Investigator Group – Intuitive XR, Brandenburg University of Technology Cottbus-Senftenberg,
Cottbus, Germany

**Introduction:** Neuroadaptive technology provides a promising path to enhancing immersive extended reality (XR) experiences by dynamically tuning multisensory feedback to user preferences. This study introduces a novel system employing reinforcement learning (RL) to adapt haptic rendering in XR environments based on user feedback derived either explicitly from user ratings or implicitly from neural signals measured via Electroencephalography (EEG).

**Methods:** Participants interacted with virtual objects in a VR environment and rated their experience using a traditional questionnaire while their EEG data were recorded. Then, in two RL conditions, an RL agent tried to tune the haptics to the user — learning either on the rewards from explicit ratings, or on implicit neural signals decoded from EEG.

**Results:** The neural decoder achieved a mean F1 score of 0.8, supporting informative yet noisy classification. Exploratory analyses revealed instability in the RL agent's behavior in both explicit and implicit feedback conditions.

**Discussion:** A limited number of interaction steps likely constrained exploration and contributed to convergence instability. Revisiting the interaction design to support more frequent sampling may improve robustness to EEG noise and mitigate drifts in subjective experience. By demonstrating RL-based adaptation from implicit neural signals, our proof-of-concept is a step towards seamless, low-friction personalization in XR.

KEYWORDS

human-computer interaction, reinforcement learning, RLHF, brain-computer interface, EEG, error detection, neuroadaptive technology

## 1 Introduction

Extended Reality (XR) has the potential to create profoundly immersive and awe-inspiring experiences. However, achieving an optimal experience requires fine-tuning various settings, from brightness and field of view to haptic feedback (Ramsamy et al., 2006) and spatial audio (Potter et al., 2022). Currently, users manually adjust these

parameters to their liking through conventional menu interfaces that closely resemble traditional desktop environments.

Unfortunately, this introduces significant friction. Frequent interruptions, particularly during initial setup, can break immersion, reduce excitement, and potentially lower long-term adoption rates. Additionally, conventional settings menus may carry a higher cost than just disrupting the immediate experience; they reposition the user into a known, age-old computing paradigm that is entirely disconnected from the immersive nature of XR. This disconnect makes personalization feel like a chore rather than a seamless and intuitive part of the experience.

Given these challenges, we set out to develop a method that effectively personalizes XR experiences while minimizing manual configuration and preserving immersion. One promising approach is to leverage Reinforcement Learning (RL), empowering an autonomous system to learn user preferences over time (Kaufmann et al., 2023). However, this presents its own obstacles, such as the need for human-provided labels and the challenge of balancing automation with user control.

One solution lies in obtaining implicit feedback from the user through neural and physiological data (Zander et al., 2016). Instead of relying on explicit user input, neural signals can serve as a real-time indicator of a user's preferences, engagement, and immersion. Here, the term 'neuroadaptive technology' is used to mark the shift from 'direct control' brain-computer interfaces (BCI) to implicit adaptation (Zander et al., 2016; Krol and Zander, 2022).

We introduce 'Neuroadaptive Haptics', a novel neuroadaptive system for a tailor-made, multisensory XR experience. 'Neuroadaptive Haptics' is an interactive system that integrates real-time neural and physiological data to dynamically modify haptics in XR environments. Our system leverages the output from a BCI as a reward signal for RL. In this paper, we applied our system to tune haptic parameters of a virtual reality (VR) system. We tested whether the system was able to dynamically adjust the VR settings to optimize users' haptic experience over time without requiring manual interventions.

# 2 Related work

Our research draws inspiration from neuroscience and engineering work on BCIs, specifically neuroadaptive technology. In order to situate our research, we provide a background on haptic experiences in XR.

## 2.1 Haptic experiences in VR

Haptic feedback in VR has been shown to be a key component in creating a realistic user experience. In fact, for a long time now, researchers have argued that attaining haptic realism is the next grand challenge in virtual reality (Brooks, 1999). In most use cases, the goal of haptic devices is to render (i.e., generate artificial touch sensations through devices) realistic sensations that mimic the sensory experience a user would normally expect when interacting with the real world. For instance, multisensory haptic renderings can combine vibrotactile feedback with force feedback, rendered via exoskeletons or electrical muscle stimulation (EMS), to simulate not only the sensation of touch but also the resistance and

rigidity of objects (Lopes et al., 2015). Additionally, other sensory cues–such as carefully synchronized auditory feedback or subliminal EMS signals–are increasingly employed to enhance haptic illusions (Cho et al., 2024; Takahashi et al., 2024). These methods can trick the brain into perceiving properties like texture, weight, or even the subtleties of material composition, by engaging multiple sensory pathways simultaneously. However, complex haptic interactions are still error-prone. The synchrony of sensory information relies on the quality of motion tracking and the accuracy of feedback presentation, and incongruous temporal feedback during object interaction may occur due to technical reasons.

At its core, this relies on predictive coding mechanisms underlying sensory integration–a framework in which our brains leverage foundational models (Pouget et al., 2013), originally established through interactions with the physical world, to interpret sensory information in both real and virtual environments (see next section). However, as evidenced by the significant technological leaps in each new hardware release, next-generation VR technology still exhibits frequent glitches and sensory mismatches, especially in key moments of multisensory integration.

## 2.2 Sampling the predictive coding mechanisms in the brain

The brain is frequently conceptualized as a predictive system that continuously generates foundational models of the environment to infer the causes of sensory input (Rao and Ballard, 1999; Friston, 2010); Clark (2013). In this framework, perception emerges from an iterative process in which predictions are compared with incoming sensory data, and discrepancies–known as prediction errors–drive model updates. These processes have been widely studied in sensory perception (Bastos et al., 2012; Keller and Mrsic-Flogel, 2018; Knill and Pouget, 2004), with research showing that the brain dynamically adjusts its internal representations to minimize these errors.

Prediction errors are particularly crucial in interactive and multimodal contexts, where sensorimotor integration plays a key role. Previous studies have demonstrated that the brain detects visuo-haptic mismatches in real-time, measurable as Prediction Error Negativity (PEN) in Electroencephalogram (EEG) responses. This neural signature has been targeted as a marker of error processing in VR scenarios, where prediction errors appear to correlate with disruptions in user experience and physical immersion (Gehrke et al., 2019; Singh et al., 2018; Si-mohammed et al., 2020; Gehrke et al., 2022; 2024). These findings suggest that predictive processing extends beyond passive perception and is deeply embedded in embodied cognition, where action and perception are considered to be tightly coupled.

## 2.3 Neuroadaptive technologies

Neuroadaptive systems are interactive technologies that adapt their behavior based on real-time neural or physiological activity (Hettinger et al., 2003). These systems aim to respond to internal user states by dynamically modifying the interface or the environment. A system can be considered neuroadaptive when it

includes a closed-loop in which neural or physiological signals are used not just for passive monitoring, but for actively shaping the user experience. Some examples include adaptive interfaces that change based on mental workload (Dehais et al., 2020) BCI-driven cursor control (Zander et al., 2016), as well as neurofeedback systems in learning and rehabilitation contexts (Mahmoudi et al., 2025).

Most applications to date have focused on desktop-based scenarios, where the challenges related to signal stability, real-time processing, and interface control are more manageable. Fewer studies have explored neuroadaptive approaches in XR, largely due to the difficulty of integrating physiological sensing in dynamic, multisensory environments. Still, early XR applications have demonstrated promise in areas like meditation support (Kosunen et al., 2016), exposure therapy for phobias (Weber et al., 2024), and adaptive training systems (Mark et al., 2022). However, these implementations generally rely on predefined scenarios and offer only limited autonomy to the computer, which typically cannot decide when or how to seek additional information from the user. Moving toward truly integrated neuroadaptive XR requires empowering the system to autonomously probe the user when necessary–sampling new data points to improve adaptation (Krol et al., 2020).

Looking beyond the scope of EEG-based measurements, other physiological data can be used for adaptive XR systems as well. For example, eye tracking signals have been explored as a measure for cognitive load in VR industrial training (Nasri et al., 2024), functional near-infrared spectroscopy (fNIRS) has been used to modulate scene intensity in a real-time VR horror experience (Berger et al., 2024), cardiac measures have guided relaxation content in immersive nature environments (Pratviel et al., 2024), and adaptive surface-EMG gesture decoders during VR object manipulation improved task success and reduced workload compared to static decoders (Gagné et al., 2025). Peripheral modalities (measuring the eyes, heart, or muscles) have strong potential, but their disadvantage is clear: They cannot access central nervous system information, i.e., the neural basis of cognition and stimulus processing and evaluation. Functional NIRS can do this, but has a temporal disadvantage compared to EEG, as it relies on the hemodynamic response, which lags several seconds behind the actual processing (Mehta and Parasuraman, 2013). Since we are interested in the direct neural responses of predictive processing, we relied on EEG in our study.

In this work, we configured an RL agent to autonomously sample the human-in-the-loop, aiming to find a haptic feedback configuration that the user experiences as the most consistent with their expectations. This approach moves beyond traditional neuroadaptive applications by combining passive physiological sensing with active decision-making in one interactive prototype, paving the way for more intelligent and responsive XR experiences. To tune the interaction over time, we used feedback, or labels, given directly by the human-in-the-loop, a special form of RL from human feedback (RLHF).

## 2.4 Reinforcement learning from human feedback

RLHF is a paradigm that enhances traditional reinforcement learning (RL) by incorporating human evaluative signals into the learning process (Kaufmann et al., 2023; Knox, 2011; Li et al., 2019.

Instead of relying solely on a predefined reward function, RLHF allows human feedback–either explicit, i.e., numerical ratings, scores, and rankings, or implicit, i.e., physiological signals such as EEG-based brain activity (Luo et al., 2018; Xavier Fidêncio et al., 2022)–to shape an agent's behavior dynamically. This approach is particularly useful in domains where reward functions are difficult to specify, such as in robotic interaction, as well as user experience optimization as it applies in adaptive XR environments.

Traditional RL systems require extensive exploration to learn optimal behaviors, which can be time-consuming and inefficient. RLHF mitigates this by enabling systems to leverage human expertise, reducing the sample complexity of learning tasks. For instance, prior work has demonstrated the effectiveness of human preference-based RL for training AI assistants, robotic control, and interactive game agents (Li et al., 2019).

A key benefit of computing rewards based on EEG-based brain activity in this context is that it enables a seamless and non-disruptive interaction: the system can adapt to the user without requiring them to stop and provide explicit input. In contrast, while explicit labels might offer more reliable signals, they introduce cognitive load and interrupt the immersive flow of the experience. EEG-based approaches address this limitation by allowing the system to adapt in the background, minimizing disruptions while still responding to changes in the user's internal states. Studies on EEG-based RLHF have shown that BCIs can provide real-time feedback signals that improve learning efficiency while reducing human effort (Xu et al., 2021; Xavier Fidêncio et al., 2022).

# 3 User study and methods

We set out to answer three questions: First, can we tune haptic rendering to participants' preferences using an RL agent based on human feedback? Second, is this possible through implicit labels obtained through a neural decoder? And third, are there disadvantages when relying on implicit instead of explicit labels?

To investigate these questions, we designed a user study where participants performed a pick-and-place task in VR: they had to pick up virtual objects and move them to designated locations. During object pick-up, visual, auditory, and haptic feedback were systematically varied to create different combinations of sensory cues. In an initial recording session, participants provided labels about the interaction through answering a question. This labeled data was then used to train a neural decoder of expected haptic sensations. Next, participants completed two blocks in which an RL agent tried to predict the haptic feedback participants deemed to best match their expectation. In one block, the RL agent operated on participants' explicit scores on the questionnaire, and in the other, it operated on the (implicit) output of the neural decoder.

## 3.1 Participants

14 participants (M = 29 years, SD = 5.2) were recruited from our local institution and through the online participant pool of the institute. Nine participants self-identified as women, 5 as men. All were right-handed (self-identification, no test). Participants were
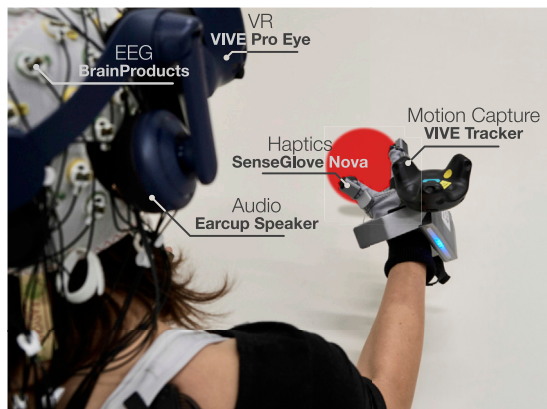
**FIGURE 1**
Experimental setup featuring a participant equipped with EEG (BrainProducts), Head-mounted display (VIVE Pro Eye), haptic feedback (SenseGlove Nova), motion capture (VIVE Tracker), and audio output (earcup speaker).

compensated with course credit or 12 € per hour of study participation. Before participating, they were informed of the nature of the experiment, recording, and anonymization procedures, and signed a consent form. The project was approved by the local ethics committee of the Department of Psychology and Ergonomics at the TU Berlin (Ethics protocol approval code: BPN_GEH_220421).

The first five participants only completed the *training* part of the experiment, which means that they did not complete the two additional blocks with the RL agent. One participant in the group who completed the blocks with the RL agent had to be excluded from any analyses concerning the agent, since they did not complete at least one of the experimental blocks due to technical issues with the EEG recording hardware. Hence, statistics for all analyses about the agent were computed for eight participants and all other analyses were computed for the full sample of 14 participants.

## 3.2 Apparatus

The experimental setup, depicted in Figure 1, comprised: (1) a VR headset with a built-in eye tracker, (2) a haptic glove with an attached motion tracker, (3) a 64-channel EEG system, and a VR capable computer (CPU: AMD Ryzen 5 5600X, GPU: AMD RADEON RX 6600 XT 8GB). To assist readers in replicating our experiment, we provide the necessary technical details, the complete source code for the VR experiment, the collected data, and the analysis scripts[1].

(1) VR. We used an HTC Vive Pro Eye 1 headset (HTC Corporation, Taiwan) to display the scene and sampled eye-tracking using SRanipal (Tobii AB, version Core SW 2.16.4.67). We replaced the stock strap of the headset with

the Vive Deluxe Audio Strap to ensure a good fit and reduce discomfort potentially caused by the EEG cap.

(2) Haptic Glove. The SenseGlove Nova V1.0 (SenseGlove, Netherlands) was used to sample finger movements and render vibro-tactile sensations. To track hand movements, an HTC Vive tracker was attached to the glove, as recommended by the manufacturer.

(3) EEG Setup. EEG data was recorded from 64 actively amplified wet electrodes using BrainAmp DC amplifiers (BrainProducts, Germany) with a high-pass filter set at 0.016Hz. Electrodes were placed according to the 10-system (Chatrian et al., 1985). One electrode was placed under the right eye to provide additional information about eye movements (vEOG). After fitting the cap, all electrodes were filled with conductive gel to ensure proper conductivity, and electrode impedance was brought below 10kΩ where possible. EEG data was recorded with a sampling rate of 250 Hz. The reference electrode was 'FCz' and the ground electrode 'AFz'.

Motion data of the hand and eye-gaze was streamed using custom scripts of 'labstreaminglayer' (LSL) (Kothe et al., 2024). Additionally, EEG data and an experiment marker stream that marked sections of the study procedure were streamed using LSL. LSL's LabRecorder was used to collect all data streams with timestamps.
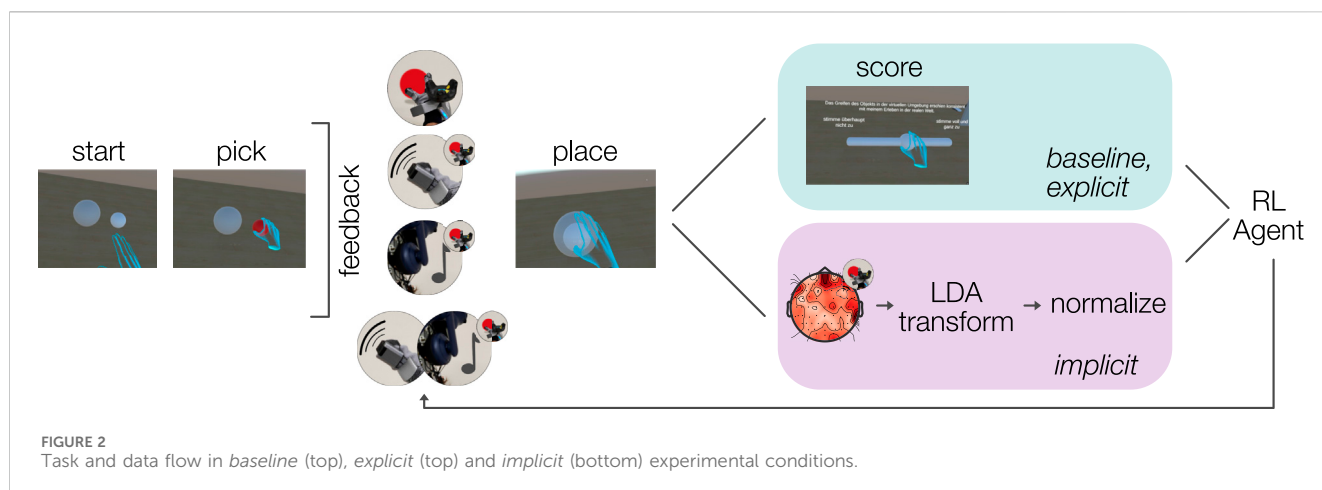
## 3.3 Task

In the task, participants were instructed to pick up an object and then place it in a target location. They were instructed to be accurate while maintaining a steady pace. The object was placed on a table in front of them and for grabbing they used the grab functionality of the haptic glove. Successful grabbing required them to use all fingers, ensuring that both the thumb and at least one other finger securely held the object. After picking up the object, the participants moved it to the center of a semi-transparent sphere that visually indicated the location of the goal. Once the object was released, participants received feedback about their placement accuracy, displayed as a numerical value in centimeters on the table, indicating the distance between the object's placement and the center of the goal sphere.

Depending on the trial condition (see procedure below), participants were then asked to rate their experience concerning the prompt "My experience in the virtual environment seemed consistent with my experiences in the real world." (translated) which was chosen from the *Multimodal Presence Scale* (Makransky et al., 2017). The anchors of the prompt were "completely disagree" and "strongly agree". To give their answer, participants could move a (continuous) slider handle by grabbing it in the same way as they were grabbing the object in the task; see Figure 2. For every trial the slider handle was reset to the center, i.e., a score of 0.5.

### 3.3.1 Interface conditions

The pick-and-place interaction was designed to simulate a multimodal interaction with visual, auditory, and haptic feedback. To this end, the following sensations were rendered:

---

1  https://lukasgehrke.github.io/neuroadaptive-xr

**FIGURE 2**
Task and data flow in *baseline* (top), *explicit* (top) and *implicit* (bottom) experimental conditions.

(1) Visual Baseline. The object changed its color from white to red when it was grabbed.
(2) Visual with Sound. Together with the color change from white to red, a sound was played at 50% volume through the Vive's earcup speakers. We used a simple 'plop' like sound with a duration of 200 m.
(3) Visual with Vibrotactile. With the color change, vibrotactile sensations were rendered at the tip of the thumb, index–and middle finger, and the back of the hand.
(4) Visual with Sound and Vibrotactile. Color change, sound, and vibrotactile feedback were rendered together.

### 3.3.2 Procedure

In total, participants completed three experimental blocks. In the first block, 140 trials had to be completed, with each interface condition being experienced 35 times. The order of the interface conditions was randomized. After every pick-and-place, the questionnaire was presented and participants had to give their score of the preceding pick-and-place interaction. These labeled data were later used to train and assess the neural decoder.

After this first block, the features for the decoder were extracted from the EEG data, and the decoder was trained. This took about 5–10 min, during which participants could rest. Next, the two experimental conditions of interest were conducted. The order in which the conditions were tested was counterbalanced across participants.

In the *explicit* condition, participants rated every interaction using the slider. The slider values were extracted as 0–1 and fed forward to the RL agent, see 3.4 for technical details on the leveraged RL implementation. The agent then selected the interface condition of the next trial. In the *implicit* condition, the question was omitted and instead the output from the decoder, after normalization to the 0–1 range, was fed to the agent. As in *explicit*, the agent then selected the interface condition of the following trial.

For both experimental conditions, we set the agent to stop after picking the same interface condition 5 times in a row, i.e., the *convergence* criterion–a threshold selected based on a series of pilot experiments. With this criterion we hoped to approximate a practical level of confidence that the agent will stabilize on a preferred policy. Furthermore, this threshold limited the overall interaction time, kept the experiment feasible for participants and

prevented fatigue while at the same time provided some robustness against noisy or inconsistent feedback, ensuring that convergence was not triggered by brief fluctuations or outlier responses in the reward signal.
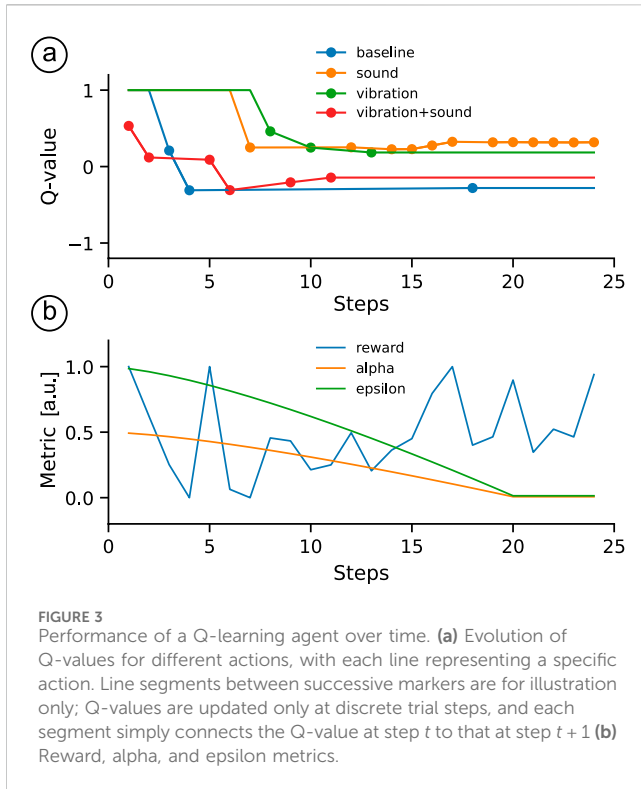
## 3.4 Reinforcement learning agent

Our RL agent was designed to select the interface condition that best matched the participant's expectation and, given that our experimental setting involved a single state, the problem reduced to a multi-armed bandit with four possible actions–essentially, a scenario where the agent repeatedly chooses from four options to maximize its reward over time. Given that our experimental setting involved a single state, the problem reduced to a multi-armed bandit with four possible actions, i.e., a scenario where the agent repeatedly chooses from four options to maximize its reward over time. In this respect, our environment was similar to that of (Porssut et al., 2022), and we adopted their validated implementation, while acknowledging that other features of the environment might not be equivalent. By following (Porssut et al., 2022), our agent also employed a combined strategy consisting of an epsilon-greedy approach and the Upper Confidence Bound (UCB) method (Auer et al., 2002), thereby balancing exploration and exploitation. This combined strategy promotes the selection of actions with high estimated Q-values while still encouraging the exploration of less-visited actions.

For our setting with $A = 4$ actions, let $\mathcal{A} = \{1, 2, 3, 4\}$ denote the set of available interface conditions. We defined the UCB value as:

$$\text{UCB}(a) = Q(a) + c\sqrt{\frac{\log_{10}(t)}{N(a)}},$$

where action $a$ was one of the four discrete interface conditions, $Q(a)$ was the current Q-value estimate for action $a$ (initially set to 1), $c$ was the exploration parameter (set to 0.25) and $N(a)$ was the number of times action $a$ had been selected up to time $t$. For actions that had not yet been chosen (i.e., $N(a) = 0$), the corresponding UCB value was set to $\infty$ to ensure they were explored. The constant $c = 0.25$ was chosen in a small suite of pilot simulations using the pure UCB algorithm under our human-feedback noise model. We swept $c \in \{0.25, 0.5, 1.0\}$ and found that 0.25 offered the most robust

FIGURE 3
Performance of a Q-learning agent over time. **(a)** Evolution of Q-values for different actions, with each line representing a specific action. Line segments between successive markers are for illustration only; Q-values are updated only at discrete trial steps, and each segment simply connects the Q-value at step $t$ to that at step $t + 1$ **(b)** Reward, alpha, and epsilon metrics.

convergence behavior. In addition to time constraints, we carried this same $c$ value forward into the hybrid model for data collection to preserve consistency and stability when layering $\varepsilon$-greedy exploration on top of UCB. This also helped minimize unexpected agent behavior during the real-time user-facing experiment.

At each time step $t \in \mathbb{N}_0$, the agent selected an action $a_t \in \mathcal{A}$ according to the following policy:

$$a_t = \begin{cases} \text{a random action chosen} \\ \text{uniformly from } \mathcal{A} & \text{with probability } \varepsilon(t), \\ \underset{a \in \mathcal{A}}{\text{argmax}}\ \text{UCB}(a) & \text{with probability } 1 - \varepsilon(t). \end{cases}$$

After executing the chosen action and observing a reward $r$, the agent updated its Q-value for that action using:

$$Q(a) \leftarrow (1 - \alpha)Q(a) + \alpha\left( r - \gamma \max_{a' \in \mathcal{A}} Q(a') \right).$$

Any action not selected on a given trial simply retained its Q-value from the previous time step. Additionally, this update rule differs from the traditional Q-learning rule, but in our tests it converged more quickly and reliably. We suspect that, due to the inherent noise in both human and neural feedback, individual Q-value estimates may fluctuate. By using $\max_{a'}Q(a')$ as a reference, the update was anchored to the best observed performance across all actions, which minimized the impact of these fluctuations and mitigated the risk of unreliable updates when an action's Q-value temporarily deviated due to noise.

Unlike the reference implementation (Porssut et al., 2022), which used an adjusted reward defined as the mode of the reward

history for action $a$, we chose to use the actual, non-adjusted reward $r$ since this yielded better performance in our pre-tests.

The initial learning rate $\alpha$ was set to 0.5, the exploration rate $\varepsilon$ was initially set to 1, and the discount factor was set to $\gamma = 0.95$. Both $\alpha$ and $\varepsilon$ decayed over time as follows:

$$\alpha(t) = \max\left( \alpha_{\min}, \alpha - \frac{\log_{10}(t + 1)}{40} \right) \quad \text{with } \alpha_{\min} = 0.001,$$

$$\varepsilon(t) = \max\left( \varepsilon_{\min}, \varepsilon - \frac{\log_{10}(t + 1)}{20} \right) \quad \text{with } \varepsilon_{\min} = 0.01.$$

This decay mechanism follows the reference implementation in (Porssut et al., 2022) and allowed the agent to gradually shift from exploration to exploitation as it gathered more information about the user's preferences. All parameters, with the exception of the UCB exploration constant $c$, were initialized to the same values as in the reference implementation (Porssut et al., 2022), as our simulations confirmed that these settings were suitable for our task. Figure 3 illustrates the agent performance for one episode.

## 3.5 Neural signal decoder

For loading, synchronizing, and pre-processing the EEG data from the 140 trials of labeled training data, we utilized the EEGLAB (Delorme and Makeig, 2004) toolbox with wrapper functions from (BeMoBIL-pipeline Klug et al., 2022) running in a MATLAB 2023b environment (The MathWorks Inc., USA).

Our goal was to design a general decoder of the expected 'haptic' sensations in VR. Therefore, we presumed the most salient neural data to be present directly following the 'haptic' event of picking up the object in our pick-and-place task. Hence, we extracted 1 s long data segments following this event and trained a binary model to score expected–unexpected sensations.

In the first step to prepare the data for decoder training, noisy (extremely large amplitude fluctuations) trials were rejected. To this end, the EEGLAB function 'autorej' was used on the EEG data, keeping the default parameters: a voltage threshold of 1,000 $\mu$V, a standard deviation cutoff of 5, a maximum rejection rate of 5% per iteration, and all channels included. This function iteratively rejects epochs with improbable data based on standard deviation thresholds, adaptively increasing the threshold when too many epochs exceed the limit.

Next, further trials were flagged and removed by detecting extreme outliers in participants' behavior. We used Tukey's method Tukey (1949), excluding values exceeding $\pm 1,5 * IQR$ (Interquartile Range) to determine trials where participants' behavior deviated significantly with respect to the placement accuracy of the target object. This resulted in the exclusion of an average 16.1 (SD = 2.9) per participant.

In line with earlier work, the EEG data was band-pass filtered to retain frequencies from 0.1 to 15 Hz via Fast Fourier Transform (FFT) prior to feature extraction (Gehrke et al., 2022; Zander et al., 2016).

### 3.5.1 Features
To obtain a robust decoder with many samples for training, we reduced the classification problem to a binary situation. Using a

median split on the questionnaire scores, we grouped all trials below the median into a *mismatching expectation* class and all trials above into *matching expectation*, resulting in a balanced dataset with two classes of, at most, 70 trials per class (minus the rejected trials).

The filtered data of all channels was then segmented into 12 epochs of interest of 50 m size from 0 to 600 m following the grab event. The samples in each 50 m segments were then aggregated using the mean, resulting in a matrix of 64 channels X 12 aggregated time windows. Next, baseline correction was performed by subtracting the first time window, i.e. 0–50 m after the grab. the event was used to subtract the physical noise of the vibrotactile stimulation from the data. This post-event window was chosen to capture and remove early physical artifacts from the vibrotactile stimulation. The first two time windows were discarded, resulting in a 64 X 10 feature matrix retained for each trial.

For each participant, a paired t-test was performed for each feature (channel and time window) to identify the most discriminative features between the *mismatching* and *matching* conditions. The absolute value t-statistics were then sorted and stored. The top 100 were then used in a grid-search on the number of features to use for decoding. To this end, the decoder's accuracy was assessed at 10 to 100 features, increasing in steps of 5. The number of features resulting in a model with the highest accuracy was saved for real-time application. These models were always fit on 80% of the data (80–20 train-test split using a 5-fold cross-validation scheme). To assess the decoder's performance, the following performance metrics were calculated: Accuracy, F1 score, and ROC.

We used a linear discriminant analysis (LDA) with shrinkage regularization (automatic shrinkage using the Ledoit-Wolf lemma (Ledoit and Wolf, 2004)) using the implementations from scikit-learn (Pedregosa et al., 2012). To normalize LDA scores during real-time application to the 0–1 range, we extracted the LDA scores of the 20% test data held out during model fitting. Then, for min-max normalization, the 5th and 95th percentiles were taken from that distribution and stored for normalization of single trials during real-time application. After normalization, any value that exceeded either 0 or 1, was then set to 0 or 1, respectively. This allowed us to retrieve 0–1 values from single trials from the binary representation of matching vs mismatching feedback expectations.
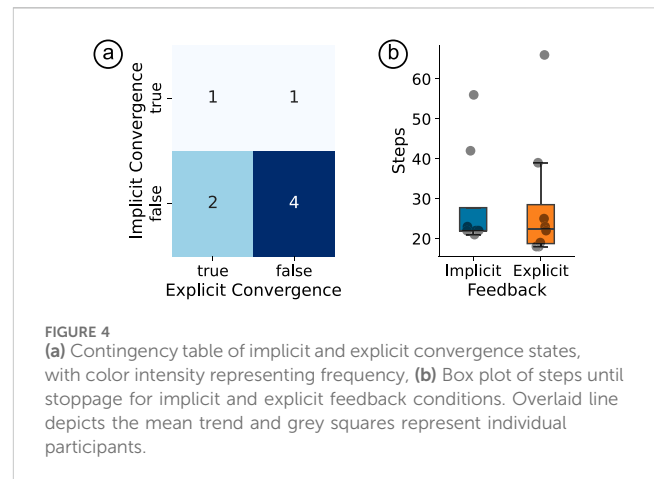
### 3.5.2 Real-time application

During real-time application, the EEG data was buffered for one second following a grab. The data was band-pass filtered analogously to the training data from 0.1 to 15 Hz. Next, the features from the best performing model were extracted and, using that model, transformed to an LDA score. Using the min-max anchors, the score was normalized to the 0–1 range and sent to an LSL stream in order to be fetched by the RL agent.

## 3.6 Hypotheses and statistical testing

To answer whether haptic rendering can be tuned using an RL agent, we inspected whether the agent arrived at the *true* label. The true label was operationalized by selecting the interface condition with the highest mean score in block 1, i.e., the training data.

We hypothesized that the RL agent would require the same number of steps until convergence in both implicit and explicit



**FIGURE 4**
**(a)** Contingency table of implicit and explicit convergence states, with color intensity representing frequency, **(b)** Box plot of steps until stoppage for implicit and explicit feedback conditions. Overlaid line depicts the mean trend and grey squares represent individual participants.
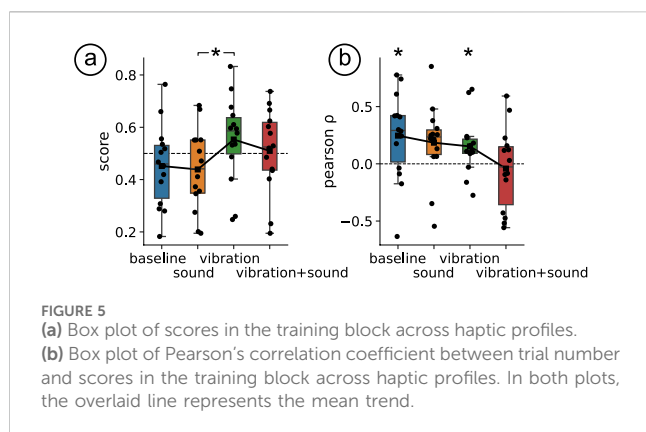
feedback conditions. Hence, the system would perform identically, irrespective of the origin of the feedback. As a reminder, the agent finished picking when it chose the same label 5 times in a row. To test the hypothesis, two one-sided *ttest* (TOST) were conducted. We decided to set the equivalence bounds to five steps, meaning that within ±5 steps difference, we determined the systems to perform equally.

To test whether our experimental manipulation was effective a linear mixed effects model was fit to explain the single-trial questionnaire scores. The interface conditions were entered as a fixed effect and a random intercept was added for each participant. Hence, the model was specified as 'score ~ interface-condition + (1– participantID)' and fit using the *pymer4* package Jolly (2018). A test statistic was obtained by calculating likelihood-ratio tests comparing the full model as specified above against the null model's score ~ 1 + (1– participantID)'. All parameters were estimated by maximum likelihood estimation Pinheiro and Bates (2006). We computed post-hoc pair-wise tests for 'interface-condition' corrected for multiple comparisons (Tukey method) using the *emmeans* package (Lenth et al., 2020).

In this work, we did not formulate any *a priori* hypotheses regarding the cortical sources or specific electrode sites contributing to the EEG-based classification. Our approach was exploratory, with all channels initially considered. Sensor-level importance emerged through a data-driven feature selection process during grid search, rather than through predefined regions of interest (ROIs). However, we present ERP waveforms at electrode Cz for visual inspection, given its central location and frequent use in prior literature.

### 3.6.1 Post-hoc analysis: Participants' scoring consistency over time

In this study, we were generally interested in how an RL agent can handle the noise inherent any reward provided through a neural interface. On top of that, we noticed that participants scoring behavior also exhibited some noise (over time). To address this source of noise in subjective scoring behavior a correlation analysis between time on task and subjective scores was conducted. The Pearson correlation coefficient (Pearson, 1895) between trial number and corresponding score was computed as a summary statistic per participant. Next, we tested whether the coefficients differed from 0 using *ttest* (Student, 1908) on the group level.

**FIGURE 5**
**(a)** Box plot of scores in the training block across haptic profiles.
**(b)** Box plot of Pearson's correlation coefficient between trial number and scores in the training block across haptic profiles. In both plots, the overlaid line represents the mean trend.

# 4 Results

The contingency table in Figure 4a summarizes the cases of convergence when either implicit or explicit labels were used. We observed that three out of eight times, the agent converged on the 'correct' feedback when using explicit user scores as rewards. On the other hand, the agent converged correctly for 2/8 participants when using implicit rewards. For two participants, the agent converged correctly using explicit rewards, but incorrectly using implicit rewards. For four participants, neither reward origin resulted in the agent converging.

b. However, the equivalence bounds were not met (lower bound: $t_{(7)} = 1.22$, $p = .26$, upper bound: $t_{(7)} = 1.22$, $p = .26$). Hence, we observed evidence that there actually is a difference in convergence time. We noticed that for either reward source, several runs converged after about 25–30 picks by the agent.

The mean difference in the number of steps to convergence between implicit and explicit reward sources was 0 (SD = 11.6), see Figure 4b. However, the TOST procedure did not yield significance (lower bound: $t_{(7)} = 1.22$, $p = .26$; upper bound: $t_{(7)} = 1.22$, $p = .26$), meaning that statistical equivalence within the predefined bounds could not be established.

## 4.1 Task validation

The four interface conditions significantly influenced participants' ratings of how consistent the virtual experience felt compared to their real-world expectations ($\chi^2_{(3)} = 9.6$, $p = .02$). Among the conditions, vibration feedback was rated as most consistent with real-world experience, while sound feedback received the lowest ratings. A follow-up comparison revealed that vibration was rated significantly higher than sound ($t_{(45.2)} = 2.8$, $p = .037$); see Figure 5a.

We found that in both interface conditions 'baseline' ($t_{(14)} = 2.41$, $p = .031$), and 'vibration' ($t_{(14)} = 2.28$, $p = .04$) participants changed their scoring behavior over time, see Figure 5B. In these conditions, their scoring became increasingly more positive with time on task (Pearson $\rho$ for 'baseline': M = 0.25, SD = 0.38, and 'vibration': M = 0.15, SD = 0.25).

## 4.2 Decoder performance

Visual inspection of the amplitudes at electrode Cz revealed an increase in the difference between *high* and *low* scored trials towards the later stages of the 100–500 m window used for classification, see Figure 6a. Indeed, the most frequently leveraged time windows as determined by the grid-search were the last three windows, i.e., 400–450, 450–500, and 500–550 m following the grab event (see Figure 6b). In terms of sensors, the five most leveraged channels were TP10, AFz, AF8, T8, and C6, see Figure 6c.

The automatic selection of the number of features used for classification for each participant resulted in an average of 40.7 (SD = 25) features being picked by the procedure. The classifier cross-validation resulted in a mean accuracy of 0.7 (SD = 0.06) and a mean F1 score of 0.8 (SD = 0.06), see Figure 6d for the mean, as well as individual participants' ROC.
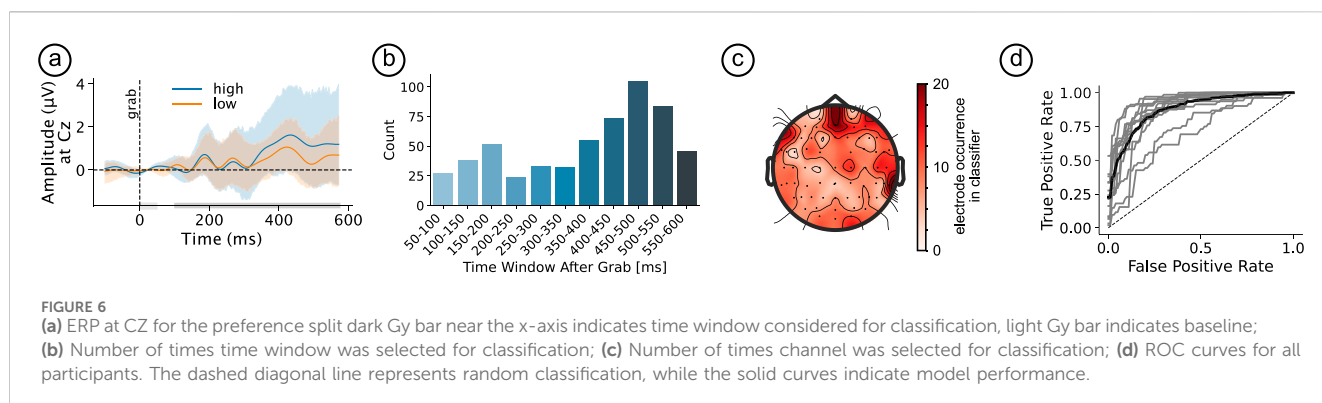
# 5 Discussion

In this paper, we set out to answer three questions: (1) Can we tune haptic rendering to participants' preferences using an RL agent based on human feedback? (2) Is this possible through implicit labels obtained through a neural decoder? And (3) are there disadvantages when relying on implicit instead of explicit labels?

We investigated these questions by building a novel, proof-of-concept, neuroadaptive system comprising an LDA-based EEG decoder and a UCB-based RL agent. The system was designed to automatically select the multisensory haptic experience for the human-in-the-loop in each following trial. We found the EEG decoder to operate at satisfactory levels (Mean F1 score of ~ .8). However, in our real-time application scenario, we observed considerably poor performance with the agent converging to the 'correct' solution in around 25% of cases across the two different reward sources. Due to the small number of participants completing the full RL protocol ($n = 8$), and the limited number of converged runs, the statistical confidence and generalizability of our findings are constrained. Future work should aim to expand the sample size to improve reliability and enable more robust conclusions about system performance and user variability. Here, we discuss our exploratory findings and provide a roadmap for future work.

## 5.1 RL agents learning from (noisy) human feedback

Regarding the first question–whether we can tune haptic rendering to participants' preferences using an RL agent based on human feedback–we found the performance of our prototype to be hindered by noisy rewards, particularly those derived from the EEG decoder. In human-in-the-loop settings, where rewards can be both noisy and non-stationary, effective exploration becomes critical. This contrasts with conventional RL scenarios, where reward functions are typically stable and well-defined.

In our study, the RL agent took approximately 25–30 steps per episode to converge on one of the four haptic feedback conditions. While this relatively short duration aligns with the simplicity of our

FIGURE 6
**(a)** ERP at CZ for the preference split dark Gy bar near the x-axis indicates time window considered for classification, light Gy bar indicates baseline; **(b)** Number of times time window was selected for classification; **(c)** Number of times channel was selected for classification; **(d)** ROC curves for all participants. The dashed diagonal line represents random classification, while the solid curves indicate model performance.

VR pick-and-place task, it contrasts significantly with typical RL scenarios that often involve thousands or millions of actions (e.g., in game-playing AI or simulated robotics environments). Shorter episodes facilitate rapid convergence and manageable computational demands, making them practical for real-time neuroadaptive applications. Futhermore, they reduce cognitive and physiological fatigue, resulting in more stable neural rewards. Conversely, longer episodes, e.g., exceeding 10 min, risk significant shifts in cognitive states.

However, limited episode length inherently restricts the agent's ability to explore the action space thoroughly. In our data, with just 25–30 steps per episode, the RL agent might not have gathered sufficient experience to robustly model nuanced haptic expectations or adapt effectively to subtle reward variations over time. Indeed, the instability in convergence we observed indicates that limited step count posed a crucial bottleneck. The noise and variability in participant ratings also highlight the need for higher-frequency sampling. Future improvements in interaction design, such as brief, minimally intrusive feedback probes or passive, implicit feedback triggered by specific user behaviors, could enhance sampling frequency without disrupting user immersion.

Conversely, while longer episodes might provide richer exploration, they introduce challenges such as user fatigue and drift in subjective ratings, complicating the agent's policy learning. Future research could thus explore hybrid or hierarchical RL approaches, incorporating long-term memory or meta-learning frameworks designed to better handle limited step counts and dynamic user feedback. Dynamically adjusting episode lengths based on real-time measures of learning progress or reward consistency could further optimize the balance between exploration depth, computational efficiency, and user experience.

In our implementation, we chose to combine $\varepsilon$-greedy with UCB exploration but omitted additional noise-handling mechanisms. Both $\varepsilon$-greedy and UCB are mechanisms designed to ensure exploration, and thus one might consider using both simultaneously to be redundant. One could argue that to achieve a higher level of exploration–one of the reasons given by Porssut et al. (2022) for this dual approach–it would suffice to simply increase the UCB exploration constant $c$, thereby pushing the algorithm to explore more aggressively. However, in practice this is not equivalent. UCB's exploration bonus naturally decreases as actions are sampled, meaning that in noisy or non-stationary settings, the associated confidence intervals can

shrink too quickly, which may lead the agent to prematurely settle on suboptimal actions. By contrast, a fixed $\varepsilon$ in an $\varepsilon$-greedy strategy guarantees that even well-sampled arms are occasionally revisited, an advantage that becomes especially important when human feedback is noisy and the environment is subject to change.

Our Q-learning update deviated from the traditional rule by anchoring each step directly to $\max_{a'} Q(a')$. Although this modification was introduced to accelerate convergence under noisy human and neural rewards, it also served as an implicit regularizer: by filtering out erratic spikes or dips in individual action values, it promotes smoother learning trajectories. However, this anchoring may bias the agent toward historically high-valued actions, which could reduce exploratory behavior in environments where reward contingencies shift over time. Future work could investigate adaptive anchoring strategies or hybrid update schemes that retain robustness to noise while preserving sufficient exploration, especially in multi-state or non-stationary settings.

In this work, we decided against using a perturbed rewards mechanism in the final solution, deviating from previous implementations Porssut et al. (2022). The combination of UCB with $\varepsilon$-greedy exploration already averages out random fluctuations in the reward signal over many trials; when the noise is moderate, the RL agent's inherent averaging means that extra corrections—such as majority voting—do not significantly change the outcome. Moreover, our empirical data showed that the RL agent converged to a stable threshold even without the perturbed rewards, suggesting that this additional mechanism was redundant since its intended effect of cleaning up noisy rewards was already achieved by the standard exploration–exploitation dynamics.

## 5.2 BCI

Turning to the second question–whether tuning is possible using implicit labels from a neural decoder–we observed that the difference in the number of steps to convergence between explicit and implicit rewards was zero. However, the TOST procedure did not confirm statistical equivalence (i.e., performance was not significantly similar within the predefined bounds). This does not imply a meaningful difference between the two, but rather that the data were insufficient to establish equivalence. Across both reward

sources, several runs converged after roughly 25–30 selections, indicating overall similar but highly variable convergence behavior.

Our implementation used a simple, linear EEG classifier that was subject-specific. This choice was motivated by the implementation simplicity, the need for low-latency deployment (low camputational demands), as well as the near-optimal performance of linear models on EEG features with roughly homoscedastic class distributions (Blankertz et al., 2011; Lotte et al., 2018). Classifiers were trained on labeled data collected around 10 minutes prior to deployment. While necessary for counterbalancing, this setup introduced a risk of temporal overfitting where the model may have adapted too specifically to neural states at training, and performed less well if participants' cognitive states shifted during the session (Hosseini et al., 2020).

For a first look at interpretable signals, we visualized ERP waveforms at electrode Cz. However, this was a post-hoc decision rather than based on predefined ROIs. Future work should more systematically relate decoding features to known ERP components such as the P3 or Prediction Error Negativity (PEN). In particular, fronto-central sites like electrode FCz, which is often associated with activity originating in or near the anterior cingulate cortex (ACC), may be especially informative due to their established role in feedback evaluation. Our grid search for feature selection consistently prioritized late time windows (e.g., 400–550 m after the grab event), aligning with the temporal dynamics of evaluative ERP components like the P3 complex (Metcalfe, 2017), and suggesting that decoding relied more on later-stage cognitive appraisal processes than early sensory encoding. Understanding which brain regions contribute to decoding is especially interesting given our multisensory task, where participants evaluated haptic and multimodal stimuli, likely engaging a broad network spanning primary and secondary sensory cortices (via modality-specific thalamic inputs), posterior insula, and posterior parietal cortex (Sathian, 2016; Andersen and Buneo, 2002). The rating process itself probably involved valuation networks including the insula, dorsal anterior cingulate cortex, and prefrontal areas such as the orbitofrontal, dorsolateral, and ventromedial cortex (Menon and Uddin, 2010; Bartra et al., 2013). This distributed cortical activation may explain the absence of a single dominant spatial feature (see scalp map in Figure 6c) and supports the view that classification relied on signals distributed across multiple EEG channels.

While our linear decoder served as a baseline, more expressive models (e.g., CNNs or RNNs) could be explored. Coupled with explainable AI techniques such as saliency maps or layer-wise relevance propagation (Farahat et al., 2019; Nam et al., 2023), these could help reveal which spectral and/or spatial EEG components drive feedback decoding. This could not only improve classification but also provide deeper insight into the neural underpinnings reflecting how user experience unfolds over time.

As EEG devices become wireless (Niso et al., 2023), more compact (Kaongoen et al., 2023), and increasingly integrated into XR hardware[2], questions of wearability and long-term comfort are

moving into focus, an essential step toward real-world deployment of neuroadaptive systems. Still, key challenges remain. EEG is highly prone to artifacts from movement and muscle activity, especially in naturalistic settings (Jungnickel et al., 2019; Gramann et al., 2011; Makeig et al., 2009; Klug et al., 2022). Real-time deployment also requires low-latency pipelines, which limits the complexity of models and preprocessing. Moreover, EEG signals vary substantially within and across individuals, often requiring personalized calibration or adaptive learning approaches (Wan et al., 2021; Wu et al., 2022). Finally, data privacy is a critical concern: because neural signals are sensitive and potentially identifiable, processing EEG data locally on the XR device instead of streaming it to external servers, can help minimize privacy risks while supporting faster, more secure interaction.

## 5.3 Task and procedure

Finally, regarding the third question–are there disadvantages when relying on implicit rather than explicit labels–we observed several challenges. A primary difficulty was the variability of user-provided labels over time. As discussed above, RL algorithms typically rely on stable reward signals, yet participant ratings often fluctuated. Correlation analyses revealed gradual shifts in subjective scores in some haptic feedback conditions as the experiment progressed, specifically for the visual-only baseline and vibration. This suggests that repeated exposure influenced participant judgments and potentially introduced biases into the RL process.

A related methodological consideration was our convergence criterion: the RL agent stopped adapting after selecting the same feedback condition five consecutive times. Participants were not informed about this convergence criterion, potentially leading to unintended confusion or frustration if they perceived no clear pattern in system responses. Future studies might explicitly communicate adaptive goals or provide intermediate feedback to clarify the system's intent, helping participants form more consistent expectations.

We also observed substantial individual differences in rating distributions. Some participants showed near-binary preference structures, consistently rating one condition as highly consistent with real-world experience while rejecting others, whereas others exhibited more graded preferences, implying nuanced perception of sensory integration. This divergence complicates RL-based adaptation, as binary structures favor rapid convergence, while graded responses introduce greater noise. Note that in this study the slider was reset to the midpoint after each trial, potentially inducing a central-tendency bias. Such a bias might have dampened extreme ratings, thereby flattening the reward gradient and slowing adaptation. Consequently, our findings in fact may underestimate rather than overestimate the achievable performance of the RL agent. Furthermore, as participants repeatedly moved the slider using substantial arm movements, first to grab the handle and then to drag it, fatigue likely increased over time. This may have led to two effects: (1) reduced rating variability, amplifying the central-tendency bias, and (2) a directional bias favoring the physically more comfortable movement direction. Future iterations, leveraging a scale without a pre-placed handle, may eliminate this confound to some degree, thereby sharpen the reward signal and accelerate convergence.

2  https://galea.co

Another potential confounding factor was an anchoring effect. Depending on the haptic condition participants first experienced, subsequent ratings might have been influenced by initial exposure. Ideally, pseudo-randomizing the initial condition to balance early experiences could mitigate this effect. However, in our study, the starting condition was fully randomized.

Together, these anchoring and scale biases may have further complicated training and interpreting the neural decoder. High heterogeneity in rating distributions, ranging from bimodal to unimodal response patterns, made it challenging to generate consistent labels for classifier training. Overall, while implicit neuroadaptive rewards offer an interesting alternative to explicit ratings, human perception and neural appraisal remain dynamic, context-sensitive, and influenced by both methodological and psychological factors. Future work should explore adaptive mechanisms accounting for evolving preferences, biases, and individual variability, ultimately supporting robust and scalable neuroadaptive XR systems.

## 6 Conclusion

In this study, we presented a proof-of-concept neuroadaptive XR system using RL to adapt multisensory haptic feedback based on explicit user ratings and implicit EEG-derived feedback. Although our EEG decoder achieved satisfactory offline performance, real-time RL performance was hindered by noisy and non-stationary feedback signals, resulting in an RL agent with poor performance.

Several critical challenges emerged from our study. Short episodes allowed rapid convergence but limited the RL agent's exploration capacity, resulting in poor overall performance. Conversely, longer episodes risk cognitive fatigue and shifting user preferences. Future work should explore interaction designs that enable higher-frequency feedback sampling. Regarding EEG-based implicit rewards, potential temporal shifts in user cognitive states complicated decoder generalization. Periodic recalibration or transfer learning paradigms spanning multiple sessions or users could mitigate this. Our linear EEG decoder prioritized later ERP components. We think it to be important to systematically link EEG decoding features with established ERP (e.g., P300, PEN) and spectral components, to enhance interpretability and neuroscientific insights, a promising avenue for future research. Finally, we noted significant variability in explicit user ratings. These biases likely introduced additional noise, impeding RL adaptation.

While we believe implicit EEG-based rewards offer an interesting alternative with a high upside to using explicit labels, we encountered both approaches to be challenged by dynamic, noisy, and context-sensitive human feedback. Addressing these challenges will be key to advancing neuroadaptive XR systems that are more attuned to the nuances of human experience.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by the local ethics committee of the Department of Psychology and Ergonomics at the TU Berlin (Ethics protocol approval code: BPN_GEH_220421). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

LG: Formal Analysis, Conceptualization, Project administration, Methodology, Supervision, Data curation, Investigation, Software, Visualization, Writing – original draft, Writing – review and editing. AK: Methodology, Software, Visualization, Formal Analysis, Validation, Writing – original draft, Writing – review and editing. MK: Methodology, Supervision, Writing – review and editing. KG: Resources, Funding acquisition, Conceptualization, Writing – review and editing.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. To copy-edit author-generated text.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Andersen, R. A., and Buneo, C. A. (2002). Intentional maps in posterior parietal cortex. *Annu. Rev. Neurosci.* 25, 189–220. doi:10.1146/annurev.neuro.25.112701.142922

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* 47, 235–256. doi:10.1023/a:1013689704352

Bartra, O., McGuire, J. T., and Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76, 412–427. doi:10.1016/j.neuroimage.2013.02.063

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi:10.1016/j.neuron.2012.10.038

Berger, C., Tennent, P., Spence, J., Maior, H. A., Ramchurn, R., and Wilson, M. L. (2024). "Adapting virtual reality horror experiences using fNIRS," in *Proceedings of the fNIRS biennial meeting 2024*.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components — a tutorial. *Neuroimage* 56, 814–825. doi:10.1016/j.neuroimage.2010.06.048

Brooks, F. P. (1999). What's real about virtual reality? *IEEE Comput. Graph. Appl.* 19, 16–27. doi:10.1109/38.799723

Chatrian, G. E., Lettich, E., and Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *Am. J. EEG Technol.* 25, 83–92. doi:10.1080/00029238.1985.11080163

Cho, H., Sendhilnathan, N., Nebeling, M., Wang, T., Padmanabhan, P., Browder, J., et al. (2024). SonoHaptics: an audio-haptic cursor for gaze-based object selection in XR. Proc. 37th Annu. ACM Symposium User Interface Softw. Technol. *(New York, N. Y. U. S. A ACM)* 8, 1–19. doi:10.1145/3654777.3676384

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi:10.1017/s0140525x12000477

Dehais, F., Lafont, A., Roy, R., and Fairclough, S. (2020). A neuroergonomics approach to mental workload, engagement and human performance. *Front. Neurosci.* 14, 268. doi:10.3389/fnins.2020.00268

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi:10.1016/j.jneumeth.2003.10.009

Farahat, A., Reichert, C., Sweeney-Reed, C. M., and Hinrichs, H. (2019). Convolutional neural networks for decoding of covert attention focus and saliency maps for EEG feature visualization. *J. Neural Eng.* 16, 066010. doi:10.1088/1741-2552/ab3bb4

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi:10.1038/nrn2787

Gagné, G., Azad, A., Labbé, T., Campbell, E., Isabel, X., Scheme, E., et al. (2025). Context informed incremental learning improves myoelectric control performance in virtual reality object manipulation tasks.

Gehrke, L., Akman, S., Lopes, P., Chen, A., Singh, A. K., Chen, H.-T., et al. (2019). "Detecting visuo-haptic mismatches in virtual reality using the prediction error negativity of event-related brain potentials,", New York, New York, USA: ACM Press, CHI '19, 427 1–11. doi:10.1145/3290605.3300657

Gehrke, L., Lopes, P., Klug, M., Akman, S., and Gramann, K. (2022). Neural sources of prediction errors detect unrealistic VR interactions. *J. Neural Eng.* 19, 036002. doi:10.1088/1741-2552/ac69bc

Gehrke, L., Terfurth, L., Akman, S., and Gramann, K. (2024). Visuo-haptic prediction errors: a multimodal dataset (EEG, motion) in BIDS format indexing mismatches in haptic interaction. *Front. Neuroergonomics* 5, 1411305. doi:10.3389/fnrgo.2024.1411305

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T.-P., Lin, C.-T., et al. (2011). Cognition in action: imaging brain/body dynamics in Mobile humans. *Rev. Neurosci.* 22, 593–608. doi:10.1515/rns.2011.047

Hettinger, L. J., Branco, P., Encarnacao, L. M., and Bonato, P. (2003). Neuroadaptive technologies: applying neuroergonomics to the design of advanced interfaces. *Theor. Issues Ergon.* 4, 220–237. doi:10.1080/1463922021000020918

Hosseini, M., Powell, M., Collins, J., Callahan-Flintoft, C., Jones, W., Bowman, H., et al. (2020). I tried a bunch of things: the dangers of unexpected overfitting in classification of brain data. *Neurosci. Biobehav. Rev.* 119, 456–467. doi:10.1016/j.neubiorev.2020.09.036

Jolly, E. (2018). Pymer4: connecting R and python for linear mixed modeling. *J. Open Source Softw.* 3, 862. doi:10.21105/joss.00862

Jungnickel, E., Gehrke, L., Klug, M., and Gramann, K. (2019). "Chapter 10 - mobi—mobile brain/body imaging," in *Neuroergonomics*. Editors H. Ayaz and F. Dehais (Academic Press), 59–63.

Kaongoen, N., Choi, J., Woo Choi, J., Kwon, H., Hwang, C., Hwang, G., et al. (2023). The future of wearable EEG: a review of ear-EEG technology and its applications. *J. Neural Eng.* 20, 051002. doi:10.1088/1741-2552/acfcda

Kaufmann, T., Weng, P., Bengs, V., and Hüllermeier, E. (2023). A survey of reinforcement learning from human feedback. *arXiv*. doi:10.48550/arXiv.2312.14925

Keller, G. B., and Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron* 100, 424–435. doi:10.1016/j.neuron.2018.10.003

Klug, M., Jeung, S., Wunderlich, A., Gehrke, L., Protzak, J., Djebbara, Z., et al. (2022). The BeMoBIL pipeline for automated analyses of multimodal Mobile brain and body imaging data. *bioRxiv* 2022, 510051. doi:10.1101/2022.09.29.510051

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi:10.1016/j.tins.2004.10.007

Knox, W. B. (2011). Augmenting reinforcement learning with human feedback

Kosunen, I., Salminen, M., Järvelä, S., Ruonala, A., Ravaja, N., and Jacucci, G. (2016). "RelaWorld," in Proceedings of the 21st international conference on intelligent user interfaces *(new York, NY, USA: acm)*.

Kothe, C., Shirazi, S. Y., Stenner, T., Medine, D., Boulay, C., Grivich, M. I., et al. (2024). *The lab streaming layer for synchronized multimodal recording.* bioRxivorg Hollow, New York: Cold Spring Harbor Laboratory. doi:10.1101/2024.02.13.580071

Krol, L. R., Haselager, P., and Zander, T. O. (2020). Cognitive and affective probing: a tutorial and review of active learning for neuroadaptive technology. *J. Neural Eng.* 17, 012001. doi:10.1088/1741-2552/ab5bb5

Krol, L. R., and Zander, T. O. (2022). "Defining neuroadaptive technology: the trouble with implicit human-computer interaction," in *Current research in neuroadaptive technology* (Elsevier), 17–42.

Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88, 365–411. doi:10.1016/s0047-259x(03)00096-4

Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. (2020). Package 'Emmeans. *R. package version* 34, 216–221. doi:10.32614/CRAN.package.emmeans

Li, G., Gomez, R., Nakamura, K., and He, B. (2019). Human-centered reinforcement learning: a survey. *IEEE Trans. Hum. Mach. Syst.* 49, 337–349. doi:10.1109/thms.2019.2912447

Lopes, P., Ion, A., and Baudisch, P. (2015). "Impacto: simulating physical impact by combining tactile stimulation with electrical muscle stimulation," in *Proceedings of the 28th annual ACM symposium on user interface software and technology* (New York, NY, USA: Association for Computing Machinery), 11–19.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-Based brain–computer interfaces: a 10 year update. *J. Neural Eng.* 15, 031005. doi:10.1088/1741-2552/aab2f2

Luo, T.-J., Fan, Y.-C., Lv, J.-T., and Zhou, C.-L. (2018). "Deep reinforcement learning from error-related potentials *via* an EEG-Based brain-computer interface," in IEEE international conference on bioinformatics and biomedicine (BIBM) *(IEEE)*.

Mahmoudi, A., Khosrotabar, M., Gramann, K., Rinderknecht, S., and Sharbafi, M. A. (2025). Using passive BCI for personalization of assistive wearable devices: a proof-of-concept study. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 1. doi:10.1109/TNSRE.2025.3530154

Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. J., and Poizner, H. (2009). Linking brain, mind and behavior. *Int. J. Psychophysiol.* 73, 95–100. doi:10.1016/j.ijpsycho.2008.11.008

Makransky, G., Lilleholt, L., and Aaby, A. (2017). Development and validation of the multimodal presence scale for virtual reality environments: a confirmatory factor analysis and item response theory approach. *Comput. Hum. Behav.* 72, 276–285. doi:10.1016/j.chb.2017.02.066

Mark, J. A., Kraft, A. E., Ziegler, M. D., and Ayaz, H. (2022). Neuroadaptive training *via* fNIRS in flight simulators. *Front. Neuroergonomics* 3, 820523. doi:10.3389/fnrgo.2022.820523

Mehta, R. K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7, 889. doi:10.3389/fnhum.2013.00889

Menon, V., and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667. doi:10.1007/s00429-010-0262-0

Metcalfe, J. (2017). Learning from errors. *Annu. Rev. Psychol.* 68, 465–489. doi:10.1146/annurev-psych-010416-044022

Nam, H., Kim, J.-M., Choi, W., Bak, S., and Kam, T.-E. (2023). The effects of layer-wise relevance propagation-based feature selection for EEG classification: a comparative study on multiple datasets. *Front. Hum. Neurosci.* 17, 1205881. doi:10.3389/fnhum.2023.1205881

Nasri, M., Kosa, M., Chukoskie, L., Moghaddam, M., and Harteveld, C. (2024). "Exploring eye tracking to detect cognitive load in complex virtual reality training," in *2024 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct)* (IEEE), 51–54.

Niso, G., Romero, E., Moreau, J. T., Araujo, A., and Krol, L. R. (2023). Wireless EEG: a survey of systems and studies. *Neuroimage* 269, 119774. doi:10.1016/j.neuroimage.2022.119774

Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* 58, 240–242. doi:10.1098/rspl.1895.0041

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2012). Scikit-learn: machine learning in python. *arXiv [cs.LG]*. doi:10.48550/arXiv.1201.0490

Pinheiro, J., and Bates, D. (2006). Mixed-Effects models in S and S-PLUS. Springer Science and Business Media.

Porssut, T., Hou, Y., Blanke, O., Herbelin, B., and Boulic, R. (2022). Adapting virtual embodiment through reinforcement learning. *IEEE Trans. Vis. Comput. Graph.* 28, 3193–3205. doi:10.1109/tvcg.2021.3057797

Potter, T., Cvetković, Z., and De Sena, E. (2022). On the relative importance of visual and spatial audio rendering on VR immersion. *Front. Signal Process* 2. doi:10.3389/frsip.2022.904866

Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* 16, 1170–1178. doi:10.1038/nn.3495

Pratviel, Y., Bouny, P., and Deschodt-Arsac, V. (2024). Immersion in a relaxing virtual reality environment is associated with similar effects on stress and anxiety as heart rate variability biofeedback. *Front. Virtual Real* 5. doi:10.3389/frvir.2024.1358981

Ramsamy, P., Haffegee, A., Jamieson, R., and Alexandrov, V. (2006). "Using haptics to improve immersion in virtual environments," in *Lecture notes in computer science* (Berlin, Heidelberg: Springer Berlin Heidelberg: Lecture notes in computer science), 603–609.

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi:10.1038/4580

Sathian, K. (2016). Analysis of haptic information in the cerebral cortex. *J. Neurophysiol.* 116, 1795–1806. doi:10.1152/jn.00546.2015

Si-mohammed, H., Lopes-dias, C., Duarte, M., Jeunet, C., and Scherer, R. (2020). *Detecting system errors in virtual reality using EEG through error-related potentials*, 653–661.

Singh, A. K., Chen, H.-T., Cheng, Y.-F., King, J.-T., Ko, L.-W., Gramann, K., et al. (2018). Visual appearance modulates prediction error in virtual reality. *IEEE Access* 6, 24617–24624. doi:10.1109/access.2018.2832089

Student (1908). The probable error of a mean. *Biometrika* 6 (1), 1. doi:10.2307/2331554

Takahashi, A., Tanaka, Y., Tamhane, A., Shen, A., Teng, S.-Y., and Lopes, P. (2024). "Can a smartwatch move your fingers? Compact and practical electrical muscle stimulation in a smartwatch," in *Proceedings of the 37th annual ACM symposium on user interface software and technology* (New York, NY, USA: ACM), 1–15.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics* 5, 99–114. doi:10.2307/3001913

Wan, Z., Yang, R., Huang, M., Zeng, N., and Liu, X. (2021). A review on transfer learning in EEG signal analysis. *Neurocomputing* 421, 1–14. doi:10.1016/j.neucom.2020.09.017

Weber, R., Dash, A., and Wriessnegger, S. C. (2024). "Design of a virtual reality-based neuroadaptive system for treatment of arachnophobia," in *2024 IEEE international conference on metrology for eXtended reality, artificial intelligence and neural engineering (MetroXRAINE)* (IEEE), 255–259.

Wu, D., Xu, Y., and Lu, B.-L. (2022). Transfer learning for EEG-Based brain–computer interfaces: a review of progress made since 2016. *IEEE Trans. Cogn. Dev. Syst.* 14, 4–19. doi:10.1109/tcds.2020.3007453

Xavier Fidêncio, A., Klaes, C., and Iossifidis, I. (2022). Error-related potentials in reinforcement learning-based brain-machine interfaces. *Front. Hum. Neurosci.* 16, 806517. doi:10.3389/fnhum.2022.806517

Xu, D., Agarwal, M., Gupta, E., Fekri, F., and Sivakumar, R. (2021). Accelerating reinforcement learning using EEG-Based implicit human feedback. *Neurocomputing* 460, 139–153. doi:10.1016/j.neucom.2021.06.064

Zander, T. O., Krol, L. R., Birbaumer, N. P., and Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proc. Natl. Acad. Sci. U. S. A.* 113, 14898–14903. doi:10.1073/pnas.1605155114