Understanding Motor Learning Experiences with Personalized Multi-Haptic Feedback

Kyungyeon Lee, Arnav Dadarya, Fumeng Yang, and Jun Nishida

***Abstract*— We explored how users experience motor learning with personalized multi-haptic feedback. To this end, we implemented Haptimize, a computational optimization system for the order and type of haptic feedback to be provided (EMS, vibrotactile, or exoskeleton) by jointly modeling objective performance and subjective factors (e.g., agency, effort, and confidence). Haptimize supports two optimization modes: fully automatic optimization, which adapts feedback based solely on performance and subjective scores, and collaborative optimization, which additionally incorporates explicit user preference responses (i.e., Like/Dislike) during optimization. A user study (N=18) on a piano learning task showed that automatic optimization improved performance while collaborative optimization enhanced subjective experience, indicating complementary roles for the two modes. Compared with prior manual customization, both computational modes achieved comparable learning performance while reducing configuration time from 48.5 s to under 10 s, making multi-haptic motor learning substantially more practical.**

I. Introduction

Numerous haptic devices have been proposed to support motor learning in domains such as musical instrument training, sports skills, and rehabilitation. These span diverse actuation forms such as exoskeletons [11], vibrotactile actuators [4], and electromagnets [6]. These systems have shown beneficial impacts, including enhanced motor performance and learning efficiency [11], better control of body movements [12], and increased user engagement throughout the learning process [8]. While these haptic systems demonstrated significant improvements in motor learning performance, several studies have shown that **accounting for subjective factors**, such as agency or confidence, during learning is essential not only for performance gains but also for long-term skill retention [5]. For novice learners in particular, accounting for subjective factors can strengthen learning motivation [2], which plays a crucial role in sustaining active participation and lowering the barrier to initial learning.

As such, **customization of multiple haptic modalities** has been explored. Prior studies have demonstrated that feedback delivered through multiple haptic modalities provides richer cues that enhance user engagement [3]. Building on these findings, recent work [7] examined user-driven (manual) customization of multi-haptic feedback sequences (e.g., force feedback for earlier learning trials, tactile feedback for later trials) for piano skill learning and showed that it can improve both learning performance and subjective experience by leveraging the distinct benefits of each modality. However, despite its effectiveness in capturing individual preferences, manual customization requires time ($\approx$ 50 s) and is often burdensome [1], time-consuming [10], and prone to inaccuracies [15], as it depends heavily on users' optimization skills and subjective judgment. Furthermore, minimizing interruptions during practice is particularly important for novices in motor skill learning, as configuration overhead increases extraneous cognitive load and detracts from task-focused learning [9]. In this context, shorter computation time is critical, as it preserves continuous learning, whereas manual customization can interrupt this process.

All authors are with the Department of Computer Science, University of Maryland, College Park, MD 20742, USA. {kylee, arnav1, fmy, jun}@umd.edu

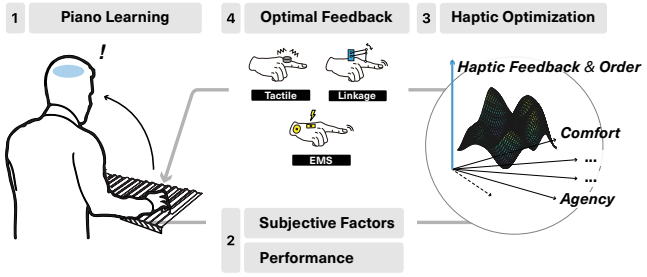


Fig. 1. Our Approach: (1) The system assists piano learning and (2) evaluates multiple aspects, including subjective factors and performance. (3) Based on these evaluations, it generates an optimal haptic sequence tailored to each user, provided through (4) an integrated multi-haptic exoskeleton.

Within this space of computational optimization, human-in-the-loop (HiL) optimization has emerged as a promising approach across HCI and robotics, integrating user preference signals into computational loops to balance objective outcomes with subjective experience. However, whether HiL provides consistent benefits over fully automated optimization in motor learning remains an open question, as user input itself imposes additional cognitive load that may disrupt learning fluency [9]. This tension motivates a systematic comparison between fully automated and HiL optimization in the multi-haptic motor learning context.

II. Our Approach: Understanding Learning Experiences with Personalized Haptic Feedback

To understand the challenges and opportunities of personalized multi-haptic feedback provision during motor skill learning, we conducted a comparative study of fully automated optimization and HiL optimization. To guide our research, we formulate the following research questions: **RQ1** asks how incorporating real-time user preference feedback (Like/Dislike) in HiL optimization impacts the motor learning experience. **RQ2** explores the trade-offs between fully

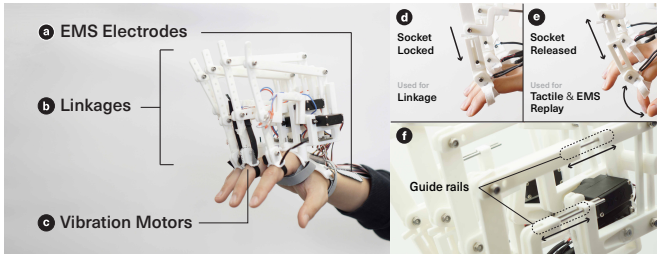


Fig. 2. Left: Multi-Haptic Exoskeleton system overview, consisting of (a) EMS, (b) linkages, and (c) vibration motors. Right: (d) linkage feedback locks sockets, (e) EMS/tactile feedback or replay releases sockets, and (f) adjustable linkages fit different hand sizes.

automated and HiL approaches, and what design implications follow for different learner profiles.

To tackle these questions, we designed and implemented Haptimize, a system that optimizes haptic feedback sequences based on motor learning (1) performance, (2) subjective factors, and (3) user preference responses (Like/Dislike), as shown in Figure 1. We chose piano learning as it requires various skill components, such as precise control over note execution, duration, and intensity, which are fundamental to motor skill learning in general.

The Haptimize system consists of three key components: **1) Integrated haptic device** that equips three different haptic modalities, including EMS, vibrotactile (tactile), and an exoskeleton (linkage), allowing for the provision of multiple haptic modalities seamlessly. **2) Computational personalization function**, optimized using individual performance and subjective factors (e.g., agency, feedback comfort), that dynamically selects the haptic feedback sequence for each learner. **3) Real-time reward mechanism**, enabling adjustments at each trial based on the user's preference responses (i.e., Like/Dislike) to the provided feedback, integrated with the personalization approach. We prepared two optimization strategies: Automatic optimization (**Auto**) adjusts the feedback sequence solely through computational personalization, without incorporating explicit user preference signals. Collaborative optimization (**Collab**) adopts a HiL design in which users provide like or dislike responses after each interaction; these preference signals are integrated with the computational personalization function to further refine the optimization process.

Through a comparative study, we provide empirical findings on personalized haptic feedback strategies and derive design implications from quantitative and qualitative data, offering learning strategies tailored to the learner's skill level.

III. Implementation

A. Multi-haptic Exoskeleton

To achieve seamless transitions between different haptic feedback types, we designed an integrated multi-haptic exoskeleton device. As shown in Figure 2, the device incorporates EMS, linkage, and tactile actuators (total weight: 312g). Three Arduino Nano ESPs operate the system as microcontrollers. First, users are equipped with Electrical Muscle Stimulation (EMS) (Figure 2a). The EMS device consists of five electrodes. The electrodes, each measuring 1 x 3 cm, are placed on the back of the user's hand, with a positive electrode on the wrist. The device provides an adjustable pulse width ranging from 0 to 300μs and operates at a stimulation frequency of 100 Hz. It consists of gate-switching devices and a DC-DC converter generating an output amplitude of 35V from AAA batteries. The average actuation latency from stimulus to playing note is 32 ms (SD = 8.3 ms). Next, users wear a hand exoskeleton (Figure 2b). This 3D-printed exoskeleton consists of four four-bar linkages, servo motors (FEETECH, FT1117M), and finger sockets. The servo motors are powered by a 6 V supply and regulated by a microcontroller. The exoskeleton is designed to actuate at a 40-degree angle, and adjusting the motor speed enables intensity control. The average actuation latency is 108 ms (SD = 19.2 ms). To enhance usability, the finger sockets are detachable, allowing users to free their fingers when receiving tactile or EMS feedback, or when replaying the melody without movement restrictions (Figure 2d and e). Furthermore, the linkages incorporate guide rails that provide a tailored fit for different hand sizes (Figure 2f). The exoskeleton integrates four linear resonant actuators (LRA) (Vybronics VG0832022D, 8 mm diameter, 3.2 mm thick, ≈ 235 Hz), one in each finger socket (Figure 2c). The actuators are embedded in 3D-printed housings and secured to the fingers with Velcro straps, ensuring both stability and easy adjustment. The LRA motors are controlled by motor drivers (Texas Instruments DRV2605L), powered by a 3 V AAA battery pack.

B. Learning Strategy Design

We prepared our approach using a reward–penalty–based optimization to dynamically select haptic modalities based on user performance and subjective factors. This class of algorithms is well-suited for our setting, as it supports sequential decision-making under uncertainty with low computational overhead, enabling real-time adaptation within a single learning session (see Appendix X for details). First, we conducted a data collection session to build a baseline computational model. The computational model learns the optimal modality-selection policy by iteratively updating the expected reward of each state-action pair: $f(s_t, a_t) \leftarrow f(s_t, a_t) + \alpha\,[r_t + \gamma \max_{a'} f(s_{t+1}, a') - f(s_t, a_t)]$, where $f(s_t, a_t)$ represents the expected reward for selecting action a_t in state s_t, α is the learning rate, r_t is the immediate reward based on performance and subjectivity, γ is the discount factor, and $\max_{a'} f(s_{t+1}, a')$ is the maximum estimated reward of the next state s_{t+1}. The state s represents the system's current context, such as the previous trials' performance scores, modality selection, and Like/Dislike feedback, while the action a denotes the selected haptic modality. The initial reward values for all modalities were set to zero. The learning rate α was set to 0.1, and the discount factor γ was set to 0.9. The exploration probability ϵ was initialized at 1.0 and decayed exponentially by a factor of 0.99 after each iteration, maintaining a minimum threshold of 0.1 to

balance exploration and exploitation. The immediate reward r_t differed by optimization mode. In *Auto*, the reward was computed solely from the objective performance score ($r_t = P$). In *Collab*, the reward combined objective performance and subjective user preference as $r_t = w_p P + w_f F$, where P denotes the performance-based reward, $F = +1$ for *Like* and $F = -1$ for *Dislike*, and $w_p = w_f = 0.5$.

IV. User Study

We conducted a two-day user study to compare learning performance and user experience between fully automated (Auto) and HiL optimization (Collab) in a piano learning task. On Day 1, participants used multiple haptic modalities (tactile, linkage, EMS), with each device evaluated separately. On Day 2, participants performed both Auto and Collab, each personalized based on Day 1 objective performance and subjective data, and Collab additionally incorporated real-time user feedback (Like/Dislike). Each session took 90 min for Day 1 and 60 min for Day 2. All procedures were IRB-approved (2330419-2), and participants received $75. The study was modeled after related work [7], which established baseline and cross-modality comparisons.

A. Task and Evaluation

Participants were asked to replay a melody, replicating the correct note, intensity, and duration. This task design was informed by prior studies on musical motor learning [17], [11], [14]. The melody consisted of six notes and was approximately 9-10 seconds long, and four types of notes were used, each corresponding to one of the following fingers: index, middle, ring, and pinky. The melody included two types of duration (short and long) and two levels of intensity (mild and strong). To ensure variety and reduce potential learning bias, the components of the melody were randomized, and participants received a different melody for each condition. Performance was evaluated on three key musical elements. To isolate execution traits from note-selection errors, duration and intensity errors were computed using only correctly played notes; incorrectly played notes were assigned the maximum error and treated as full errors in the total score.

Note Accuracy: The correctness of which notes were played compared to the given melody. In our study, we assigned four distinct notes to four fingers: D3 to the index finger, E3 to the middle finger, F3 to the ring finger, and G3 to the pinky finger. The sum of the correct number of notes across a trial was used for evaluation.

Duration Accuracy: The accuracy of the note lengths played compared to the given melody. We categorized note durations into two types: short (0.5 sec) and long (1.5 sec).

Intensity Accuracy: The accuracy of how loud the notes were played compared to the given melody. We defined two intensity levels: strong (90) and mild (50). The sum of the error values across a trial was measured from MIDI values (key press velocity, which ranges from 0 to 127).

B. Setup and Apparatus

Our setup is illustrated in Figure 4. We used a MIDI keyboard (Novation, Launchkey 61 MK3) paired with noise-canceling headphones (Sony, WH-CH720N). A tablet was positioned on the left side to administer questionnaires, and a foam hand support was provided to ensure the participants' comfort and stability during the task. For real-time user feedback, we used MIDI buttons: red indicated 'dislike', and green indicated 'like'. For software, we used Apple's Logic Pro software to record the user's piano playing, alongside AppleScript to automate the process. This ensured consistent timing between the initial feedback-based playing and the subsequent replays.

C. Conditions

We employed a within-subject study design to compare **Collab**, where participants provide real-time Like/Dislike feedback during training, and **Auto**, where the system selects feedback based solely on performance and subjective scores collected on Day 1. The order of conditions was counterbalanced across participants to minimize order effects.

D. Participants

We recruited 18 participants (8 self-identified as female, 10 as male) from our local institution. The participants had a mean age of 22.4 years (SD = 3.2). We recruited individuals with no prior piano experience or those who rated themselves as novices (at level 1-2 on a 10-point proficiency scale), as this allowed us to better isolate the effects of our evaluation without the influence of pre-existing piano skills. All participants provided written informed consent, were right-handed, and did not overlap with participants in the technical evaluation.

A power analysis was conducted to determine the sample size for the within-subjects design. A pilot study with four participants suggested a medium effect size ($f = 0.25$) for the performance outcomes. With $\alpha = 0.05$ and power set at 0.80, the analysis indicated that at least 12 participants were required. Our final sample of 18, therefore, met and exceeded this threshold.

E. Procedure

1) Calibration and Prior Training Session: Before the study, we calibrated our linkage and EMS devices for each participant (see Figure 3). For EMS, we increased the stimulation strength until the participant's finger was properly actuated or the maximum comfortable intensity was reached. We adjusted the electrical impulses by adjusting the pulse width, ranging from 1 to 300 μs. The intensity of the stimulation was adjusted accordingly. For linkage, we determined the optimal starting angle, from 1 to 10 degrees, to best accommodate each participant's hand shape. The entire calibration process took about 15 minutes.

To minimize potential bias and familiarize the participants with the study procedure, we conducted a 20-minute training session before the main study. In this session, participants practiced with a sample melody that had a similar structure to

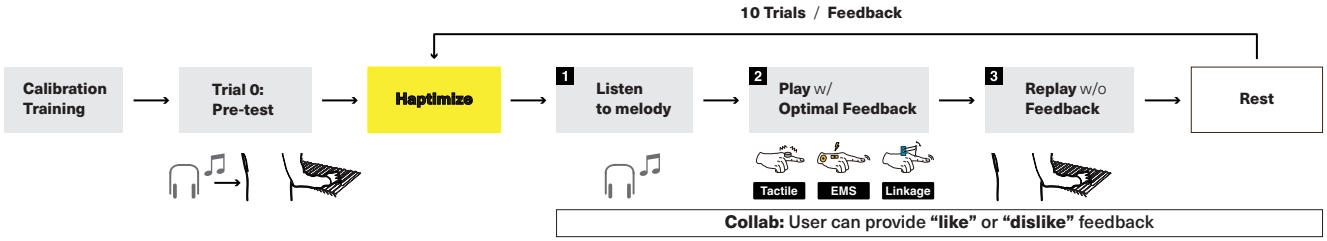


Fig. 3. Study procedure: After calibration and prior training sessions, participants perform a pre-test. Haptimize then recommends the optimal modality, followed by (1) listening to a melody, (2) playing the melody with haptic feedback, and (3) replaying the melody without feedback.

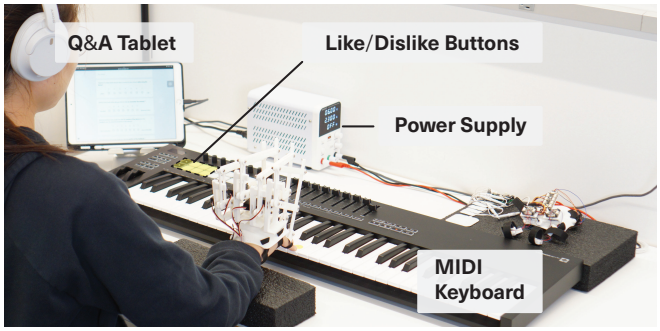


Fig. 4. Overview of the user study setup

the melodies used in the main study. They followed the same procedure as in the main study until they fully understood the task requirements. We reminded participants to remember certain aspects of the melody, such as which finger felt the feedback, how long their finger felt the feedback, and how intense the feedback was.

2) Pre-test (Trial 0): Before starting learning sessions with the feedback, participants completed a pre-test to establish their baseline performance. During this trial, participants (1) listened to a melody, then (2) replayed it without receiving any feedback.

3) Feedback and Replay: As shown in Figure 3, each condition had 10 trials (30 total), each comprising three stages: (1) **listening** to a 6-note melody (9–10 s), (2) **learning with feedback**, receiving haptic feedback to learn the melody, and (3) **replay without feedback**, replaying it while focusing on note, duration, and intensity.

4) Questionnaires: After the study, participants completed a 10-item questionnaire using 7-point Likert scales. Five items, adapted from prior work [7], captured learning experience with haptic feedback (i.e., agency, effort, confidence, and comfort), while the remaining five assessed system usability (i.e., mental demand, engagement, focus, reliability, and usefulness) (see Appendix XI for full questionnaire statements).

F. Quantitative Results

For performance analysis, we conducted a Wilcoxon signed-rank test for note accuracy and paired-sample t-tests for duration and intensity accuracy. For subjective ratings, we employed linear mixed-effects models using ARTool[16]

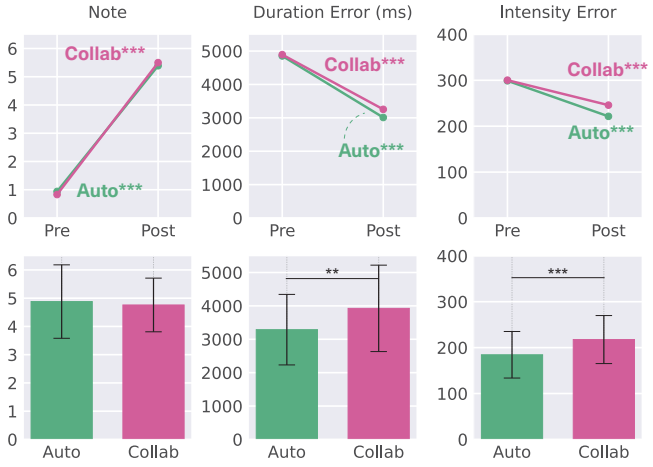


Fig. 5. (Top) Results of the pre- and post-test. (Bottom) Results of average performance scores of Auto and Collab. * indicates P < .05. ** indicates P < .01. *** indicates p < .001.

to perform non-parametric analyses. All analyses used Benjamini–Hochberg (BH) correction (FDR = 0.05).

*1) **Preference**:* Among the participants, 50% (n = 9) preferred *Collab* and 50% (n = 9) favored *Auto*.

*2) **Performance: pre- and post-test**: Pre* value was defined as performance in trial 0, while *post* was defined as the highest performance observed in later trials (trials 7–10). Figure 5 (Top) shows the results of the pre- and post-test for both conditions.

Note Accuracy in the *Auto* increased significantly from a mean of 0.94 to 5.39 ($p < 0.001, Z = -3.52, r = 0.83$). The *Collab* condition also showed a significant increase from 0.83 to 5.50 ($p < 0.001, Z = -3.61, r = 0.85$). **Duration Accuracy** in the *Auto* improved significantly from 4854.12 to 3110.50 ($p < 0.001, t(17) = -6.63, d_z = 1.56$). The *Collab* condition similarly improved from 4893.33 to 3254.16 ($p < 0.001, t(17) = -6.61, d_z = 1.56$). **Intensity Accuracy** in the *Auto* improved significantly from 298.90 to 221.44 ($p < 0.001, t(17) = -6.10, d_z = 1.44$). The *Collab* condition also showed a significant improvement from 300.25 to 245.85 ($p < 0.001, t(17) = -6.09, d_z = 1.44$).

*3) **General Performance**:* We also compared the average performance between the *Auto* and *Collab* conditions, as shown in Figure 5 (Bottom). For **Note Accuracy**, the mean score was 4.88 in the *Auto* condition and 4.76 in the *Collab*

condition, with no significant difference ($p = 0.48, Z = -0.70$). For **Duration Accuracy**, the mean error was significantly lower in the *Auto* condition (3287.82) compared to the *Collab* condition (3924.52; $p = 0.006, t(17) = -2.38, d_z = 0.56$). For **Intensity Accuracy**, the *Auto* condition also showed a significantly lower mean error (184.42) compared to the *Collab* condition (217.54; $p < 0.001, t(17) = -4.12, d_z = 0.97$).

*4) **Likert Questionnaires: System Usability**:* Subjectivity ratings related to system usability are shown in Figure 6. **Mental Demand** was higher in the *Collab* condition (M = 4.17) than in the *Auto* condition (M = 3.06), with a statistically significant difference ($p = 0.0135, t(17) = 2.75, d_z = 0.65$). **Engagement** was also higher in the *Collab* condition (M = 5.94) compared to the *Auto* condition (M = 5.33), showing a significant difference ($p = 0.0369, t(17) = 2.26, d_z = 0.53$). For **Focus**, the *Auto* condition showed a higher mean (M = 5.94) than the *Collab* condition (M = 4.89), with a statistically significant difference ($p = 0.004, t(17) = -3.32, d_z = 0.78$). **Reliability** was slightly higher in the *Auto* condition (M = 5.39) than in the *Collab* condition (M = 5.17), but the difference was not significant ($p = 0.5084$). **Usefulness** was comparable across conditions, with means of 5.33 for *Auto* and 5.44 for *Collab*, showing no significant difference ($p = 0.7421$).

*5) **Likert Questionnaires: Learning Experience**:* For learning experience, **Agency** was higher in the *Collab* condition (M = 5.89) than in the *Auto* condition (M = 5.17), with a statistically significant difference ($p = 0.0231, t(17) = 2.24, d_z = 0.53$). **Effort** ratings indicated that participants exerted less effort to remember the melody in Collab (M = 4.39) than in Auto (M = 3.56; higher scores indicate less effort), showing a significant difference ($p = 0.039, t(17) = 3.31, d_z = 0.78$). **Device Comfort** was slightly higher in the *Collab* condition (M = 4.06) than in the *Auto* condition (M = 3.78), but this difference was not significant ($p = 0.135$). **Feedback Comfort** was significantly higher in the *Collab* condition (M = 5.78) compared to the *Auto* condition (M = 4.83; $p = 0.004, t(17) = 3.29, d_z = 0.78$). **Confidence** was similar across conditions, with means of 5.28 for *Collab* and 5.56 for *Auto*, showing no significant difference ($p = 0.667$).

G. Qualitative Feedback

Like/Dislike Feedback: Engaging or Distracting? Participants expressed mixed opinions about providing Like/Dislike feedback during collaborative adaptation. Some (n = 4) found it useful for influencing the system, while others (n = 4) considered it distracting. For example, one participant remarked, *"The presence of the selection button disrupted my focus—I wanted to concentrate on playing rather than deciding whether I liked the feedback."* (P18) Relatedly, three participants indicated that interacting with the system reduced monotony: *"Playing the same melody multiple times is boring to be honest. When I have no option to click buttons, I have no choice but to receive only one type of feedback, which was very boring."* (P11).

Concerns About Personal Bias: Some participants (P4, P9, P12) expressed concerns regarding the reliability of their own preferences as a reason for favoring Auto. For instance, P4 noted, *"I also did not fully trust my own preferences, recognizing a potential bias. To perform better, I chose to follow the system's suggestions, which ultimately proved effective."*

Higher Perceived Performance with Auto: Several participants (P5, P8, P13, P17) reported that although they disliked certain feedback modalities, the system's automatic selections improved their performance. P5 noted, *"I don't like EMS, but Auto suggested EMS multiple times. I hated the experience, but I felt my performance was better."* Similarly, P13 reflected, *"At first, I preferred Collab, but considering my performance, I would switch to Auto for learning. I definitely played better with Auto."*

H. Optimized Haptic Feedback Sequences

We analyzed the distribution of modality across Auto and Collab. In Auto, EMS was selected most frequently, accounting for 45.6% of the choices, followed by Linkage at 32.2% and Tactile at 22.2%. This indicates that the automated optimization favored EMS feedback more strongly than the other modalities. In contrast, Collab showed a different pattern: Linkage was most common at 42.8%, followed by Tactile at 36.1% and EMS at 21.1%. These results suggest that when participants were involved in the optimization process, they showed a stronger preference toward Linkage and Tactile feedback, while EMS was less frequently chosen.

V. Comparison with Previous Results

We compared our results with the baseline findings reported in Hapticus [7], which examined performance with three haptic devices on the same task. The melody difficulty, experimental procedure, and hardware setup were held constant across studies, enabling a direct comparison. Both studies involved 18 participants. **Dataset:** We obtained performance data from the original authors of Hapticus [7] and used it as a reference baseline for comparison. Specifically, we analyzed data from the best-performing score set in Study 1, which employed non-computational personalization. **Findings:** We conducted Wilcoxon rank-sum tests for note accuracy and two-sample t-tests for duration and intensity accuracy, with BH correction, comparing the computational optimization methods (averaged across the *Collab* and *Auto* conditions, M_c) with the corresponding values from manual optimization reported in *Hapticus* (M_h). For each participant, we analyzed the best-performing trial within each condition. Accuracy metrics were computed following the definitions described in Sec IV-A.

Across all metrics, the computational conditions were comparable to the baseline. Note Accuracy ($M_c = 5.85, M_h = 5.50, Z = -1.79, p = 0.12$) and Duration Error ($M_c = 3613.50, M_h = 3795.40, t(17) = -0.98, p = 0.18$) showed no significant difference, while Intensity Error was significantly lower in Haptimize ($M_c = 188.80, M_h = 198.45, t(17) = -3.10, p = 0.03$).

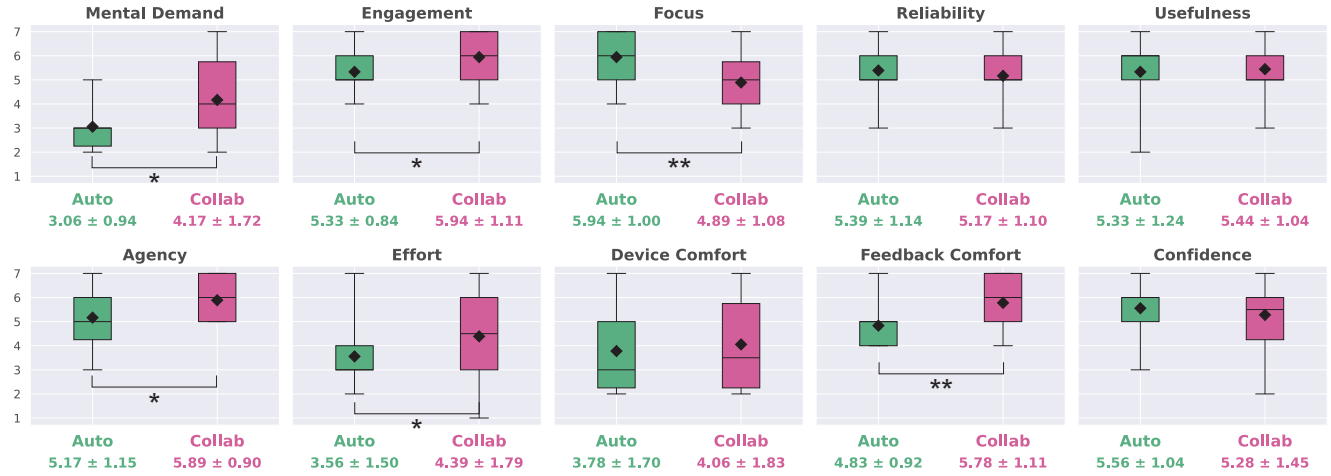


Fig. 6. Comparison of system usability (top row) and haptic learning experience (bottom row) subjective ratings between Auto (green) and Collab (pink).

Configuration Time: Auto completed haptic feedback optimization and configuration in 0.3 s on average (SD = 0.04 s), whereas Collab required 8.2 s (SD = 0.4 s). In comparison, prior manual customization [7] required 48.5 s on average (SD = 5.20 s). This near-instantaneous time is because Auto configures feedback in a single evaluation of the learned policy ($a_t = \arg\max_a f(s_t, a)$), while Collab's 8.2 s reflects waiting for the user's Like/Dislike response.

VI. Discussion

A. *The Trade-off: Subjective Factors vs. Performance*

While Collab enhanced subjective factors such as agency, engagement, and feedback comfort, it did so at the expense of objective performance on more complex metrics, including duration and intensity accuracy. In Collab, learners were required to simultaneously perform the motor task (piano playing) and an executive task (evaluating and selecting feedback preferences); this dual-task demand likely had a negative impact on performance.

Conversely, Auto functioned by "offloading" the decision-making process to the system. This reduction in cognitive load allowed participants to maintain higher focus and allocate their attentional resources entirely to the motor learning. This aligns with prior work suggesting that automation improves efficiency in motor learning by reducing task-related workload [13]. Therefore, our results highlight distinct roles for each approach: while human-in-the-loop designs are essential for fostering motivation and agency, automating the decision process can effectively reduce the burden of choice to maximize performance in high-precision motor tasks.

B. *Balancing Performance and User Experience*

One notable finding is the tradeoff between objective performance and user experience. Auto was more frequently selected EMS, leading to greater performance improvements, but participants perceived the feedback as more intrusive, resulting in lower comfort and agency.

In contrast, although Collab yielded smaller performance improvements, it significantly improved the learning experience, resulting in greater feedback comfort and a stronger sense of agency, factors associated with sustained motivation in prior work [2].

These findings suggest that different optimization strategies may be appropriate depending on the training objective. Auto may be better suited for maximizing performance improvement, whereas Collab may be preferable in learning environments that prioritize user comfort, perceived agency, and sustained engagement, particularly for novice learners.

C. *Comparison to Baseline*

Compared to the baseline established in Hapticus [7], Haptimize achieved comparable performance with a shorter configuration time. Our results show that Haptimize achieves these efficiency gains without sacrificing performance, providing empirical evidence that computational personalization can support motor skill learning by reducing the decision and exploration overhead associated with identifying the most suitable haptic device. These findings highlight the potential of multi-haptic systems combined with computational personalization to optimize motor learning processes.

VII. Design Implications

Based on our findings, we outline design implications for future multi-haptic feedback systems.

A. *Novice: Collaborative Haptic Optimization*

Collab could benefit **novice users** who require engagement and comfort to maintain learning motivation at an early stage of skill acquisition (Figure 7a). For example, in a beginner piano lesson, a novice learner could provide Like/Dislike feedback after each trial, guiding the system toward preferred haptic modalities that feel comfortable and engaging. This combination of agency and comfort can sustain motivation throughout prolonged practice sessions, supporting gradual progress toward a target skill level.

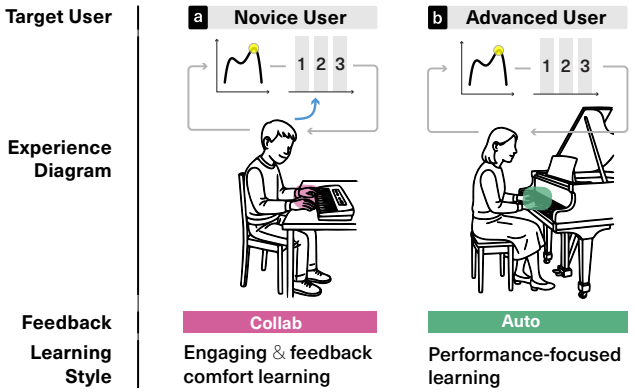


Fig. 7. Design implications: Two learning strategies based on user skill level and learning scenario.

B. Advanced: Automatic Haptic Optimization

Auto mode would be particularly useful for **advanced users** who have mastered basic motor skills and seek to refine performance toward a higher level of expertise (Figure 7b). At this stage, learners must practice efficiently within a limited time, prioritizing uninterrupted execution and sustained focus on high-level performance factors. By automatically selecting optimal haptic modalities in under one second (vs. ≈48.5 s for manual customization and ≈8.2 s for *Collab*), Auto allows the learner to remain fully focused on the motor task, thereby sustaining concentration and accelerating performance gains. For instance, an intermediate pianist refining a technique or a resident surgeon practicing a procedure under time constraints could benefit from this seamless, performance-oriented feedback delivery without the burden of manual configuration.

VIII. Limitations and Future Work

Our evaluation primarily relied on post-trial 7-point Likert scales. Future work could complement these subjective ratings with objective measures, such as physiological signals (e.g., eye tracking), to more directly capture users' experiential states and provide richer optimization signals.

Building on such richer measurements, our current approach focuses solely on optimizing the order of haptic feedback. Future work will extend the optimization space by incorporating additional haptic parameters, including feedback intensity, duration, and frequency, enabling more fine-grained and personalized assistance.

Finally, our reward-penalty-based algorithm could be extended to more advanced optimization paradigms. While it performed well in our current setup, future work could explore richer reward formulations that better balance objective performance and subjective experience. More sophisticated approaches, such as contextual bandits, Deep Q-learning, model-based reinforcement learning, Bayesian Optimization, or CMA-ES, may better capture complex user preferences, improve sample efficiency, and support long-term adaptation while maintaining low-latency interaction.

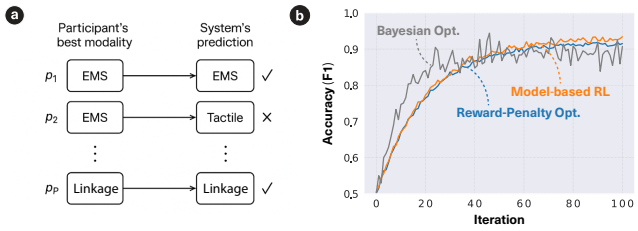


Fig. 8. (a) Accuracy calculation: each participant's self-reported best modality is compared with the system's predicted modality. (b) Convergence of optimization methods across 100 iterations.

IX. Conclusion

Motor learning experiences under personalized multi-haptic feedback differ systematically depending on how users are involved in the optimization. A user study (N=18) showed that collaborative optimization (Collab) enhanced agency and feedback comfort, while fully automatic optimization (Auto) improved duration and intensity accuracy and focus, indicating complementary strengths suited to different learner profiles. Based on these findings, we outlined design implications that match optimization strategies to learner profiles and training contexts. This comparison was enabled by Haptimize, a computational personalization system that jointly models performance and subjective factors. Haptimize achieved learning performance comparable to prior manual customization while reducing the time required for haptic feedback optimization and configuration from 48.5 s to 0.3 s (Auto) or 8.2 s (Collab), making multi-haptic motor learning significantly more practical. Together, this work contributes empirical understanding of how optimization modes shape the learning experience, practical design guidance for different learner profiles, and a personalization system that makes such experience-centered comparisons feasible.

APPENDIX

X. Technical Evaluation

a) Objective: To implement a computational optimization system for haptic feedback, we evaluated three methods used in HCI: a reward–penalty–based optimization, model-based RL, and Bayesian optimization. All methods were feasible for our design, achieving over 85% accuracy.

b) Setup: Data were collected from 10 participants in a separate session independent of the main user study, including performance and subjective factor ratings for each modality, after which participants were asked to indicate their preferred modality (Figure 8a). Using both subjective and performance data, we implemented the three methods to predict each participant's best modality. Each method was assessed in terms of average prediction accuracy, computation time, and suitability for our interactive setting. We evaluated accuracy using the F1 score by comparing the model-predicted optimal feedback type with the actual feedback type selected by each participant. The task was treated as a categorical label-matching problem, with F1 and

accuracy used to quantify the extent to which each method aligned with participants' choices.

c) Results: Figure 8(b) shows that all three optimization methods achieved high final performance, indicating their feasibility for our design. Notably, all methods reached stable convergence within approximately 40 iterations, with accuracy exceeding 85%, suggesting that the training budget was sufficient within the limited number of iterations.

Our reward–penalty–based algorithm converged stably, reaching an F1 score of approximately 0.92 with low variance, achieving accuracy comparable to Bayesian optimization and model-based RL. Given its strong accuracy, suitability for real-time interactive environments, and low computational overhead, we adopt the reward–penalty–based algorithm as the optimization method for Haptimize.

While the chosen method and hyperparameters (e.g., learning rate and exploration schedule) could be further optimized through additional experiments, our goal was to identify settings appropriate for a real-time interactive context, and they demonstrated stable and effective performance.

XI. Likert Questionnaire Statements

Agency: "To what extent did you feel in control of your actions while using the device?" (1: I did not, 7: I did).

Effort: "How much effort did you have to exert to remember the melody?" (1: Max effort, 7: No effort).

Device Comfort: "How would you rate the comfort of the device (not the feedback, but the physical property of the device)?" (1: Not comfortable, 7: Very comfortable).

Confidence: "How confident were you in replaying notes?" (1: Not confident, 7: Completely confident).

Feedback Comfort: "How would you rate the comfort of the feedback?" (1: Not comfortable, 7: Very comfortable).

Mental Demand: "How was the task mentally demanding?" (1: Very low, 7: Very high).

Engagement: "How engaging was the task?" (1: Not engaging, 7: Very engaging).

Focus: "To what extent did the system help you stay focused throughout the task?" (1: Not focused, 7: Very focused).

Reliability: "How reliable do you find this system?" (1: Not reliable, 7: Very reliable).

Usefulness: "How useful was the automatic suggestion for learning the melody?" (1: Not useful, 7: Very useful).

Acknowledgment

This work was supported by the JST Moonshot R&D Program (JPMJMS2013), the JST PRESTO Program (JPMJPR25I6), and UMD Immersive Media Design program.

References

[1] DiMicco, J.M., Mann, N.: User Research to Inform Product Design: Turning Failure into Small Successes. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. pp. 872–879. ACM, San Jose California USA (May 2016). https://doi.org/10.1145/2851581.2851606

[2] Duarte, J.E., Reinkensmeyer, D.J.: Effects of robotically modulating kinematic variability on motor skill learning and motivation. Journal of Neurophysiology **113**(7), 2682–2691 (Apr 2015). https://doi.org/10.1152/jn.00163.2014

[3] Fujita, H., Hosoi, J., Ban, Y., Warisawa, S.: Presenting Tingling Aftereffects Using Vibro-Thermal Feedback to Enhance Impact Sensation in Virtual Reality. In: 2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR). pp. 505–515. IEEE, Saint Malo, France (Mar 2025). https://doi.org/10.1109/VR59515.2025.00074

[4] Gemicioglu, T., Hopper, E., Dwivedi, B., Kulkarni, R., Bhandarkar, A., Rajan, P., Eng, N., Ramanujam, A., Ramey, C., Gilliland, S.M., Mason, C., Seim, C., Starner, T.: Passive Haptic Rehearsal for Augmented Piano Learning in the Wild. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **8**(4), 1–26 (Nov 2024). https://doi.org/10.1145/3699748, https://dl.acm.org/doi/10.1145/3699748

[5] Kao, D.L.: Plateaus and the curve of learning in motor skill. Psychological Monographs **49**(3), i–94 (1937). https://doi.org/10.1037/h0093530

[6] Lee, K., Vaichalkar, J., Dadarya, A., Chang, W., Kikumoto, A., Nishida, J.: FIXical I/O: Exploring the Effects of Real-time Error Sensing and Physical Intervention on Finger-based Motor Sequence Learning. Proceedings of the CHI Conference on Human Factors in Computing Systems (2026). https://doi.org/https://doi.org/10.1145/3772318.3790432

[7] Lee, K., Yang, D., Singh, K., Nishida, J.: Hapticus: Exploring the Effects of Haptic Feedback and its Customization on Motor Skill Learning: Tactile, Haptic, and Somatosensory Approaches. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. ACM, Yokohama Japan (May 2025). https://doi.org/10.1145/3706598.3713821

[8] Manjunatha, H., Pareek, S., Memar, A.H., Kesavadas, T., Esfahani, E.T.: Effect of Haptic Assistance Strategy on Mental Engagement in Fine Motor Tasks. Journal of Medical Robotics Research **05**(01n02), 2041004 (Mar 2020). https://doi.org/10.1142/S2424905X20410044

[9] Mark, G., Gudith, D., Klocke, U.: The cost of interrupted work: More speed and stress. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 107–110. ACM, Florence Italy (Apr 2008). https://doi.org/10.1145/1357054.1357072

[10] McGrenere, J., Baecker, R.M., Booth, K.S.: An evaluation of a multiple interface design solution for bloated software. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 164–170. ACM, Minneapolis Minnesota USA (Apr 2002). https://doi.org/10.1145/503376.503406

[11] Nishida, J., Tanaka, Y., Nith, R., Lopes, P.: DigituSync: A Dual-User Passive Exoskeleton Glove That Adaptively Shares Hand Gestures. In: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. pp. 1–12. ACM, Bend OR USA (Oct 2022). https://doi.org/10.1145/3526113.3545630

[12] Nith, R., Serfaty, J., Shatzkin, S.G., Shen, A., Lopes, P.: JumpMod: Haptic Backpack that Modifies Users' Perceived Jump. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–15. ACM, Hamburg Germany (Apr 2023). https://doi.org/10.1145/3544548.3580764

[13] Reinkensmeyer, D.J., Patton, J.L.: Can Robots Help the Learning of Skilled Actions? Exercise and Sport Sciences Reviews **37**(1), 43–51 (Jan 2009). https://doi.org/10.1097/JES.0b013e3181912108

[14] Rogers, K., Röhlig, A., Weing, M., Gugenheimer, J., Könings, B., Klepsch, M., Schaub, F., Rukzio, E., Seufert, T., Weber, M.: P.I.A.N.O.: Faster Piano Learning with Interactive Projection. In: Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces. pp. 149–158. ACM, Dresden Germany (Nov 2014). https://doi.org/10.1145/2669485.2669514

[15] Vázquez-Ingelmo, A., García-Peñalvo, F.J., Therón, R.: Tailored information dashboards: A systematic mapping of the literature. In: Proceedings of the XX International Conference on Human Computer Interaction. pp. 1–8. ACM, Donostia Gipuzkoa Spain (Jun 2019). https://doi.org/10.1145/3335595.3335628

[16] Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 143–146. ACM, Vancouver BC Canada (May 2011). https://doi.org/10.1145/1978942.1978963

[17] Yuksel, B.F., Oleson, K.B., Harrison, L., Peck, E.M., Afergan, D., Chang, R., Jacob, R.J.: Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty Based on Brain State. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 5372–5384. ACM, San Jose California USA (May 2016). https://doi.org/10.1145/2858036.2858388