Human-in-the-Loop Optimization of Perceived Realism of Multi-Modal Haptic Rendering Under Conflicting Sensory Cues

Harun Tolasa, *Graduate Student Member, IEEE*, Bilal Catkin, and Volkan Patoglu, *Member, IEEE*

***Abstract*—During haptic rendering, a visual display and a haptic interface are commonly utilized together to elicit multi-sensory perception of a virtual object, through a combination and integration of force-related and movement-related cues. In this study, we explore visual-haptic cue integration during multi-modal haptic rendering under conflicting cues and propose a systematic means to determine the optimal visual scaling for haptic manipulation that maximizes the perceived realism of spring rendering for a given haptic interface. We show that the parameters affecting visual-haptic congruency can be effectively optimized through a qualitative feedback-based human-in-the-loop (HiL) optimization to ensure a consistently high rating of perceived realism. Accordingly, the multi-modal perception of users can be successfully enhanced by solely modulating the visual feedback without altering the haptic feedback, to make virtual environments feel stiffer or more compliant, significantly extending the range of perceived stiffness levels for a haptic interface. We extend our results to a group of individuals to capture the multi-dimensional psychometric field that characterizes the cumulative effect of feedback modalities utilized during sensory cue integration under conflicts. Our results not only provide reliable estimates of just noticeable difference thresholds for stiffness with and without visual scaling but also capture all the prominent features of sensory cue integration, indicating weights that are proportional to the congruency level of manipulated visual signals. Overall, preference-based HiL optimization excels as a systematic and efficient method of studying multi-modal perception under conflicts.**

***Index Terms*—Haptic rendering, multi-modal perception under conflicts, perceived realism, preference-based human-in-the-loop optimization, sensory integration, visual-haptic congruency.**

I. Introduction

HUMANS receive real-world sensory cues through various feedback channels and combine and/or integrate them to form a robust perceptual model of the world. Sensory signals that are complementary (hence, non-redundant) are combined, while redundant sensory signals are integrated to form a coherent multi-sensory percept [1]. Bayesian models are commonly utilized to model how these sensory signals are (partially) integrated, providing insights into the neural mechanisms involved in perceptual decision-making [2]. There exists strong evidence in the literature that neural mechanisms over a population of neurons implement *sensory integration* similar to the maximum likelihood estimation (MLE) [1]. During sensory integration, a percept is obtained by a weighted linear combination of redundant sensory signals; if the weights are associated with signals according to their (perceived) reliability, then the most reliable percept is achieved [1], [3]; otherwise, the result is considered suboptimal in terms of MLE. Moreover, as the reliability of a signal is reduced, its weight in integration has been shown to decrease, possibly leading to a change of the dominant sensory modality that has the largest contribution in the percept [4], [5], [6]. Furthermore, as the discrepancy among the signals becomes large, the reliability of the multi-sensory estimate may become less than that of an unimodal estimate, leading the discrepant source to be *vetoed*, instead of being integrated [1], [5], [7], [8].

Since independent sensory cues are required for optimal MLE-like integration behavior, its applicability for modeling conflicting cues is limited. In particular, while the MLE model is likely to stay valid for small and hard-to-detect discrepancies along the dimension of interest, this model of sensory integration fails to account for the breaking down of cross-modal interactions, when the information provided by each modality is highly conflicting [9], [10].

Several extensions of the MLE model exist in the literature that can account for the partial integration of cues across a wide range of inter-modal discrepancies and stimulus conditions [11], [12]. For instance, [11] utilizes prior knowledge about the correspondence between multi-modal cues when determining the degree of integration, while the causal inference model [12] considers possible causes of the underlying sensory events to enable partial integration.

Haptic perception often involves the fusion of complementary *force-related* cues and *movement-related* cues. Moreover, visual feedback commonly accompanies the haptic perception of movement, providing *additional* movement-related cues known to dominate other modalities under many circumstances [13], [14]. During the integration of visual and haptic sensory inputs, the percept has been shown to depend on each cue [1], [5], while

Received 16 December 2023; revised 1 July 2024 and 1 December 2024; accepted 14 January 2025. Date of publication 30 January 2025; date of current version 20 June 2025. This work was supported by TUBITAK under Grant 120N523 and Grant 23AG003. This article was recommended for publication by Associate Editor A. Moscatelli and Editor-in-Chief D. Prattichizzo upon evaluation of the reviewers' comments. *(Corresponding author: Volkan Patoglu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Sabancı University Research Ethics Council under Application No. FENS-2024-10, and performed in line with the Declaration of Helsinki.

The authors are with the Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul 34956, Türkiye (e-mail: harun.tolasa@sabanciuniv.edu; bilalcatkin@sabanciuniv.edu; volkan.patoglu@sabanciuniv.edu).

Digital Object Identifier 10.1109/TOH.2025.3535416

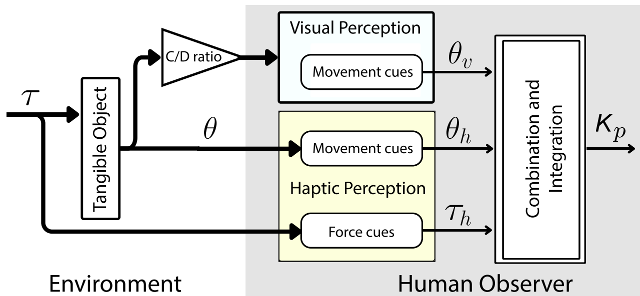

Fig. 1. A schematic representation of visual-haptic sensory integration and combination to form a coherent compliance percept. In the figure, τ denotes the torque resulting from the rotation θ of the object. Symbol θ_h represents the haptic perception of movement cues, while τ_h denotes the haptic perception of force-related cues. The visual cues due to the scaled movement displayed on the monitor are denoted by θ_v, while the scaling factor is captured by the C/D ratio. Finally, K_p denotes the perceived stiffness formed by (partial) integration and combination of multi-modal cues.

the integration process may be suboptimal from a maximum reliability perspective [15], [16].

Haptic rendering is the bilateral process that makes the dynamics of computationally mediated virtual environments (VEs) apparent to a human user through a haptic interface. Haptic interfaces are commonly complemented with a visual display to improve immersion levels. Haptic rendering aims to maximize the perceived realism of the VEs by ensuring the perceived similarity of the rendering with respect to the reference model; however, high-fidelity rendering is restricted to a limited range of VEs due to underlying hardware limitations, such as the limited force output or resolution.

During a typical haptic rendering task, sensory information provided by the haptic interface and the visual display include both complementary and redundant cues. In softness perception, movement-related percepts have been found to be formed through the integration of cues provided by both haptic and visual modalities [15], as depicted in Fig. 1. Furthermore, it is possible to induce a controlled discrepancy between haptic and visual movement cues by introducing a scaling factor, called control-to-display (C/D) ratio, to the visual feedback provided during haptic rendering.

Haptic perception of a virtually mediated object under visual-haptic congruency has been of considerable interest, since it has been demonstrated that visual cues can override haptic feedback in displacement-related tasks, indicating the potential of visual manipulation to improve haptic rendering experiences [13]. Such approaches are commonly utilized for *pseudo-haptics* [17]. In the literature, it has been demonstrated that perceived compliance/stiffness can be increased by scaling the visual displacements [13], [18], [19], [20], [21], [22]. Similar approaches have also been applied to other haptic rendering tasks, such as rendering weight [14] and surface roughness [23], [24].

Overall, it is possible to capitalize on the multi-modal and redundant nature of the movement-related cues to enhance the *perceived* rendering range of any haptic interface by manipulating the visual cues. However, it remains an open challenge to determine the proper level of visual scaling to achieve the most realistic haptic rendering for a user. This goal necessitates an understanding of the psychometric model of multi-modal haptic rendering under conflicting sensory cues.

In this study, we explore the visual-haptic cue integration and propose a systematic means to determine the optimal visual scaling during haptic manipulation that maximizes the perceived realism of multi-modal spring rendering. Our approach is rooted in sample-efficient human-in-the-loop (HiL) optimization, where the rendering parameters are iteratively updated based on participants' qualitative feedback.

We extend our results to a group of individuals to capture the underlying multi-dimensional psychometric field that characterizes the cumulative effects of feedback modalities utilized during sensory integration under conflicting cues. Our results not only provide reliable estimates of just noticeable difference (JND) thresholds for stiffness under visual scaling but also capture several prominent features of sensory integration.

Overall, we demonstrate that the HiL optimization approach allows for the determination of appropriate haptic-visual parameters. This ensures a consistently high rating of perceived realism of multi-modal rendering. Additionally, it provides a systematic and efficient means to study multi-modal sensory integration under conflicting cues.

II. Related Work

The HiL setting where participants provide feedback in each trial is commonly employed by standard methods of classical psychophysics, such as the method of constant stimuli and the method of limits. However, these studies are not well-suited for evaluating stimuli with more than one dimension because the number of trials grows exponentially with the number of dimensions and the number of points per dimension [25]. Accordingly, several adaptive techniques, predominantly based on Bayesian methods, have been developed to achieve similar accuracy with classical psychophysics methods while using fewer trials [26].

Bayesian optimization is a global optimization approach commonly employed in studies involving humans due to its sample efficiency [27]. Bayesian optimization approaches have been used for psychophysics studies [28], [29], [30], [31], [32], [33], [34], [35], [36], as well as for optimization of assistive robotic devices [37], [38], [39], [40], for which the evaluation of optimization metrics is costly or the number of trials is constrained by human involvement. Bayesian optimization methods can be loosely categorized as parametric and non-parametric approaches.

A. Parametric Approaches

The well-known adaptive methods, such as QUEST [28], and Psi [29], rely on the assumption that a parametric model for the psychometric function consistent with Weber's law exists. These parametric approaches assume that the stimulus varies on only one dimension and utilize a Bayesian update strategy to achieve sample efficiency. However, the extension of these methods to multi-dimensional stimuli is limited, as they evaluate any additional dimensions independently resulting in an inefficient search strategy over one-dimensional slices of the multi-dimensional psychometric field.

QUEST+ [31] and Psi-marginal [30] are more general parametric approaches that support multi-dimensional models, but they still require the parametric form of the psychometric field to be specified *a priori*. On the other hand, the extension of one-dimensional psychometric curves to multi-dimensional psychometric fields is not straightforward, as there may exist nonlinear interactions between the additional variables and saturation behavior at low and high intensity levels [41].

Furthermore, since all parametric methods strongly depend on the model introduced prior to data collection, the conclusions that can be drawn from these approaches become limited if the data violate the underlying assumptions of the selected model.

B. Non-Parametric Approaches

Non-parametric approaches have been introduced to remove strong assumptions about the shape of the multi-dimensional psychometric fields, by modeling the psychometric function using a general approximator, such as a stochastic process. In most cases, a Gaussian process (GP) serves as a sample-efficient non-parametric model for complex functions [34], [35]. Furthermore, non-parametric approaches replace dense sampling with efficient active learning schemes, significantly improving the applicability of these methods for psychometric studies with multivariate stimulus settings [32], [33], [34], [35], [36].

The classical implementation of GP-based Bayesian optimization has been developed for quantitative metrics and is not directly applicable to qualitative evaluations. On the other hand, psychophysical studies are dominantly based on "yes/no" type classification tasks and most HiL optimization studies rely on qualitative comparative feedback from users to enhance the reliability of subjective metrics [42].

GP-based Bayesian optimization approaches have been extended to active learning with classifications instead of quantitative measurements [43], [44], [45]. For instance, GPs for binary classifications with "yes/no" type feedback have been introduced in [43], while this qualitative feedback-based method has been extended to ordinal classifications and pairwise preferences in [45]. These GP-based Bayesian optimization approaches have been applied to psychophysics [34], [35], [41], [46], [47] and HiL optimization applications [48], [49], [50], [51].

Fewer studies use preference-driven optimization for haptic rendering [52] and texture generation [53]. Catkin et al. [52] proposed preference-based HiL optimization for spring and friction rendering, showing its effectiveness in capturing user preferences and customizing parameters to maximize perceived realism. They demonstrated that HiL optimization provides an efficient way to study the effects of haptic parameters on perceived realism, even in high-dimensional spaces.

In this study, we utilize a non-parametric Bayesian optimization approach, as this selection allows us to model the multi-dimensional psychometric field between the rendering parameters and perceived realism without assuming a fixed parametric form. Our work is built upon the existing GP-based HiL optimization approach presented in [48], [52], since GP-based models are multi-dimensional by default and are flexible enough to model correct saturation behaviors.

While previous non-parametric psychophysical studies used either binary classifications [34], [41] or pairwise preferences [35], [47] as the qualitative feedback from the users, in this study, we extend this methodology to use a wider range of classifications to get more information from the user in each trial, such that we can build a more general and sample-efficient method. Our study significantly extends and refines these approaches to suit multi-modal haptic rendering under visual-haptic incongruency and provides insight into multi-modal sensory integration by capturing the underlying psychometric field and perception thresholds. The characterized psychometric field models the cumulative effect of feedback modalities on perceived realism during sensory integration of conflicting cues.

III. Qualitative Feedback-Based HiL Optimization

The HiL optimization aims to learn a quantified relationship between the rendering parameters and user perception through a GP-based latent function. The GP-based latent function models users' perception without making strong parametric assumptions. As users provide more qualitative feedback, such as classifications and comparisons, the GP-based latent function is updated according to Bayes' theorem, to capture their perception more accurately. During the HiL experiments, the GP-based latent function is used within a Bayesian optimization framework to efficiently learn the relationship by employing informed sampling techniques.

After posterior GP models of all users are trained, a generalizable latent perception model can be constructed by statistically averaging these posterior GP models, and a probabilistic relationship between the rendering parameters and corresponding qualitative feedback outcomes can be derived from the averaged posterior GP-based latent function to extract psychophysical thresholds of interest.

While a detailed mathematical model of the HiL optimization method based on qualitative feedback is presented in Appendix A, an overview can be provided as follows:

1) Latent Function Modeling: During the HiL optimization, subjects form a perception of the stiffness K_{p1} and K_{p2} from two different visual-haptic renderings and $K_{p_{ref}}$ from the reference rendering. Users are asked to evaluate the stiffness of renderings based on their perceived similarity with respect to the reference. Our experiments collect two ordinal classifications, q_{o_1} and q_{o_2}, and one pairwise preference, q_p, at each iteration. Users provide ordinal classification based on the similarity between a rendering and the reference, and they give a pairwise comparison between two renderings based on which one is more similar to the reference. Although pairwise comparisons provide no extra information when subjects classify parameters into separate categories, pairwise preferences are useful to capture small differences if two parameters are classified in the same category.

We train a GP-based latent function $f(x)$ to learn the relationship between rendering parameters $x = (K, \text{C/D ratio})$, and users' perceived similarity by using the collected qualitative feedback data D, where K is the rendered stiffness and C/D ratio parametrizes the visual scaling. The latent function is modeled such that higher perceived similarity results in higher

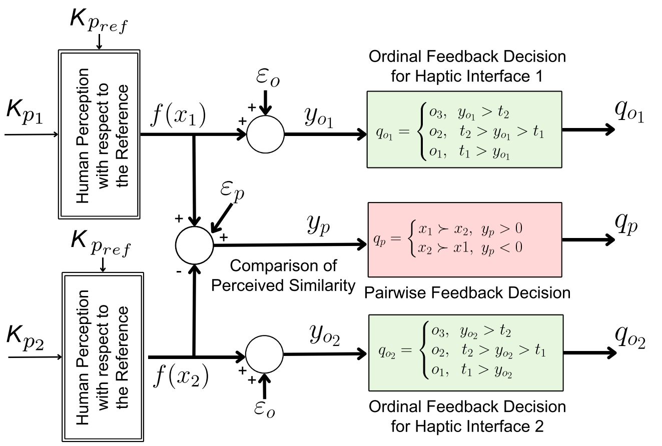

Fig. 2. Human decision model for qualitative feedback.

latent function scores. The prior probability of the latent function $P(f(x))$ is modeled using a normalized GP model.

We know that human decisions are not perfectly consistent; hence, this inconsistency is modeled by white noise interference to the decision process, as depicted with noises ϵ_o and ϵ_p in Fig. 2. The white noise aims to capture the limitations of human sensitivity and any random distractions that may take place during the experiment. Given that humans are typically more consistent while making pairwise preferences compared to ordinal classifications [50], distinct white noise parameters are used for pairwise preferences and ordinal classifications.

The users' ordinal classification decisions are modeled according to the threshold values between where the noise-interfered version of latent scores y_{o1} and y_{o2} fall. The binary classification used in [41] is a special case of ordinal classification with two categories separated by a threshold at zero. We preferred to apply three ordinal categories to enforce more reward to renderings perceived similar to reference and more penalty to renderings with high perceived distinctiveness. Similarly, the pairwise preference is modeled according to the noise-interfered version of the difference between latent function scores. When the noise-added difference is above zero, the user perceives the stiffness of the first rendering K_{p1} to be more similar to reference rendering $K_{p_{ref}}$.

2) Posterior Latent Function and Bayesian Inference: As the user provides more qualitative feedback, the latent function is updated according to Bayes' theorem

$$P(f(x)|D) \propto P(D|f(x))P(f(x)) \tag{1}$$

allowing it to capture the user's perception more precisely.

The aim of acquiring the posterior distribution of $f(x)$ is to enable predictions of the latent function scores $f(x_*)$ for any feasible arbitrary rendering parameter x_*. However, since the probability of qualitative feedback is not Gaussian, the posterior distribution of $f(x_*)$ is not analytically tractable using the posterior distribution of $f(x)$. Accordingly, the Laplace method is used to approximate the posterior distribution of $f(x)$ as a Gaussian distribution such that the posterior distribution $f(x_*)$ is also inferred as a Gaussian. Utilizing this commonly adopted method for posterior approximation, it is possible to make predictions for Bayesian optimization and extend the probabilistic derivations to capture psychophysical estimations based on human perception [44], [48], [49], [50].

Algorithm 1: Pseudo-code for HiL Optimization with Qualitative Feedback.

initiate S: Parameter space, $f \sim GP(\mu_0, \sigma_0)$: GP prior, M: Space-filling iteration number, N: Total iteration number
1: **for** $i = 1, 2, \ldots, N$ **do**
2:   **if** $i \leq M$ **then**
3:     Randomly select two distinct points from S
4:   **else**
5:     Select two parameters $x_{i_{1,2}} := \text{argmax}_{x \in S}\,(\alpha_{UCB}(x)_{i-1})$
6:   Observe qualitative feedback ($q_{o_{i_1}}$, $q_{o_{i_2}}$ and q_{p_i})
7:   Update posterior distribution of latent function model f

3) Bayesian Optimization: The Bayesian optimization uses the GP model of the latent function as defined in the previous subsection. In each iteration, a query consisting of a two-parameter set is sampled by using the Upper Confidence Bound (UCB) acquisition function

$$\alpha_{UCB}(x_*) = \mu(x_*) + c\,\sigma(x_*) \tag{2}$$

where $\mu(x_*)$ denotes the mean of $f(x_*)$, $\sigma(x_*)$, is the standard deviation of the posterior probability distribution of $f(x_*)$, and c is a constant hyper-parameter used to define the exploration/exploitation ratio. As c increases (decreases) the weight of the standard deviation increases (decreases) thus, the sampling algorithm focuses more on exploration (exploitation). To sample from different regions, we also impose a condition during the selection of the second parameter set, such that the prior covariance should be smaller than 0.2.

Algorithm 1 presents the basic steps at each iteration: The first M iterations are conducted to explore the search space via space-filling methods (Lines 1–3); i.e., random uniform sampling. Then, for the next $N - M$ iterations, the algorithm suggests two parameter sets using the most promising points according to the acquisition function (Line 5). Next, the suggested parameter sets are used in the visual-haptic rendering with the left and right knobs. After the subject tries the rendered parameters, qualitative feedback regarding the trial is transmitted to the algorithm (Line 6). Lastly, the algorithm updates its GP posterior according to the qualitative feedback data (Line 7). We use the GP update procedure for the parameter inference as in [49].

4) Aggregating Gaussian Process and Extracting Psychophysical Thresholds: After the data collection, the trained posterior models are used to create an averaged posterior GP model to form a general perception model. Then, this model is used to infer the classification and comparison decisions of users throughout the bimodal stimuli space.

a) Averaged Gaussian Process Posterior: The averaged posterior GP model is computed from the individual posterior GPs of the multi-modal experiment. We treat each participant's posterior GP results as independent and identically distributed measurements over the bimodal space and use them to obtain the average perceived similarity posterior model as follows

$$\bar{f}_{*|D_{tot}} \sim GP\left(E(\bar{f}_{*|D_{tot}}), Var(\bar{f}_{*|D_{tot}})\right)$$
$$\text{with} E(\bar{f}_{*|D_{tot}}) = \frac{1}{n}\sum_{s=0}^{n} E(f_{s*|D_s})$$
$$Var(\bar{f}_{*|D_{tot}}) = \frac{1}{n^2}\sum_{s=0}^{n} Var(f_{s*|D_s}) \tag{3}$$

where n denotes the number of subjects, while D_{tot} represents qualitative feedback data collected from all participants, D_s, represents individual qualitative feedback data, and f_s represents individual perceived similarity scores with respect to the reference.

b) Just Noticeable Difference Thresholds: Utilizing the averaged GP posterior model, one can estimate the probability of a participant preferring a rendered parameter set instead of the reference parameter set. The probability model quantifies the probability of each parameter set being as effective as the reference, in terms of perceived similarity to the reference. Let $\bar{f}_{*|D_{tot}}$ be the averaged posterior distribution of latent function f for arbitrary parameter set x_*, where $\bar{f}_{*|D_{tot}}$ is calculated according to (3). Similarly, let $\bar{f}_{ref|D_{tot}}$ be the averaged posterior distribution of latent function for reference parameter x_{ref}. Then, the probability of a participant preferring the arbitrary parameter x_* instead of reference can be calculated as

$$P(x_* \succ x_{ref}|D_{tot}) = \Phi\left(\frac{E(\bar{f}_{*|q_{tot}} - \bar{f}_{ref|D_{tot}})}{\sqrt{Var(\bar{f}_{*|D_{tot}} - \bar{f}_{ref|D_{tot}}) + c_p^2}}\right) \tag{4}$$

where $E(\bar{f}_{*|D_{tot}} - \bar{f}_{ref|D_{tot}})$ is the difference between the mean values of the posterior distributions, $Var(\bar{f}_{*|D_{tot}} - \bar{f}_{ref|D_{tot}})$ denotes the variance of the difference between $\bar{f}_{*|D_{tot}}$ and $\bar{f}_{ref|D_{tot}}$, and c_p is the coefficient used in defining the standard deviation of the white noise interference in (11).

Preferring the rendered parameter set over the reference indicates that a participant either assigns a higher perceived similarity score to the rendered parameter set compared to the reference or is unable to distinguish between the rendered parameter set and the reference; hence, randomly selects one of the options. If the probabilities of selecting the rendered parameter set and the reference are equal when the participants cannot differentiate between the two options, and if we assume that the participants always prefer the reference when they notice a difference between the two choices, then utilizing the Bayesian inference, the probability of participants preferring the reference by detecting the difference can be computed as $1 - 2P(x_* \succ x_{ref}|q_{tot})$.

To select a confidence interval where participants prefer the reference by detecting the difference with respect to 50% probability, one can select the JND threshold at 25%. Accordingly, when the parameter set falls below this threshold, the parameter set can be assumed to have a noticeable difference with a probability over 50%.

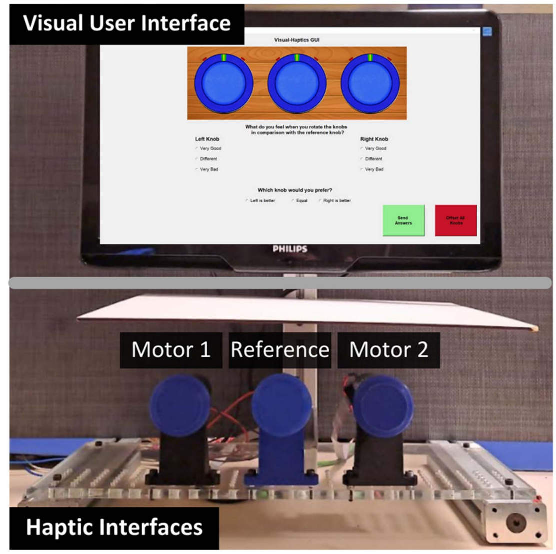

Fig. 3. Three identical haptic interfaces and a visual display.

IV. Human-in-the-Loop Optimization Experiments

A. Participants

Twelve participants (9 males and 3 females) with an average age of 25.91 $\pm$ 1.16 years participated in this study. Among the participants, only one person was left-handed. No participant had any known sensory-motor disability. Before the experiments, all participants signed an informed consent form approved by the Institutional Review Board of Sabanci University (Protocol No: FENS-2024-10). None of the participants had significant prior experience with haptic interfaces and psychophysical studies.

B. Apparatus

The experiment setup, shown in Fig. 3, consists of three identical haptic interfaces and a visual display. Equal-sized knob-shaped 3D-printed handles were used as the end-effector of the haptic interfaces. The visual display was utilized for the visual-haptic congruency experiments, such that the visual feedback of all knob rotations was made available on the monitor, while the view of the haptic interfaces was kept hidden from the participants.

Each haptic interface consists of a direct-drive Maxon RE40 brushed DC motor equipped with a 1024 count/rev encoder and driven in the current mode with a MaxPos amplifier through an EtherCAT interface. Open-loop impedance control was implemented in real-time at 1 kHz through the Matlab RealTime environment.

A torsional spring with stiffness value K was rendered according to Hooke's law, $\tau = K\,\theta$, where τ denotes the reference torque resulting from the rotation of the end-effector θ. The stiffness range of $[0.1, 0.4]$ Nm/rad was utilized in the experiments and mapped to the normalized range of $[0, 1]$ during optimization. The reference stiffness was 0.2 N-m/rad, corresponding to the normalized stiffness value of 0.5.

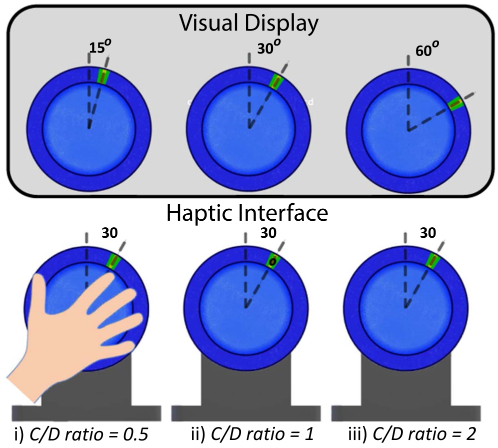

Fig. 4. Knob rotations of the haptic interfaces and corresponding displayed visual rotations for the different C/D ratios.

C. Control-to-Display Ratio and Visual/Haptic Stiffness

The perceived stiffness of virtual torsional springs is studied by controlling the visual scaling factor, called control-to-display (C/D) ratio [14] of the knob rotations. The method involves applying a visual scaling to the actual knob rotation of the haptic interface and allowing visual feedback only from the monitor, as depicted in Fig. 4.

If the C/D ratio is less (more) than one, then the visualized movement is attenuated (amplified), requiring the user to rotate the knob more (less) to observe the same amount of rotation on the screen compared to higher (lower) C/D ratios.

The C/D ratio range of $[0.5, 2]$ is considered, such that two times visual amplification or reduction can be provided for a given knob rotation. During the optimization, C/D ratios are normalized to $[0, 1]$. The inverse of the following mapping is applied to normalize the C/D ratios: $C/D_{ratio} = 2^{(2\alpha-1)}$, where α denotes the normalized C/D ratio.

Consistent with the C/D ratio definition, we define *visual stiffness* K_v as the idealized case when only the visual movement cues are used with haptic force cues to form a stiffness perception, such that $K_v \triangleq \Delta\tau_h/\Delta\theta_v$. Similarly, we also define *haptic stiffness* K_h as the unimodal case when only the haptic movement cues are used with haptic force cues to form a stiffness perception, such that $K_h \triangleq \Delta\tau_h/\Delta\theta_h$.

D. Hypotheses

We have designed an experiment to test the validity of the following hypotheses.

H1 The perception of users can be successfully manipulated by changing the visual modality without altering the rendered stiffness parameter, to make VEs feel stiffer or more compliant. It is possible to increase the perceived compliance (stiffness) via amplification (attenuation) of the visual motion feedback.

H2 The visual-haptic incongruency limits the range of visual scaling for which the perception of users can be manipulated with high perceived realism. To determine these limits, the JND thresholds for stiffness and C/D ratio can be estimated through the HiL optimization experiments.

H3 Movement-related cues from the haptics modality and visual cues are (partially) integrated through a weighted linear combination of redundant sensory cues. Accordingly, the perceived stiffness is formed as a linear combination of visual stiffness K_v and haptic stiffness K_h values. Furthermore, as the incongruency level increases, the contribution of visual cues and their weights in the stiffness perception decreases.

E. Experimental Procedure

1) Setup and Overview: All volunteers participated in the experiment through a single sitting and used their dominant hands. To minimize the effects of auditory cues, participants wore headsets playing pink noise. During the experiment, participants were instructed that the knobs control the visual display and asked to interact with the three knobs in a similar manner. Participants were allowed to explore these systems as they preferred. To ensure that the visual cues were only provided from the display, all three haptic interfaces were covered throughout the experiment.

During all sessions, the middle knob was used as the reference rendering that maintained the same stiffness value $K_{ref} = 0.2$ Nm/rad. The visual feedback for the reference (middle) knob corresponded exactly to the physical rotation of the haptic interface (C/D ratio = 1), ensuring that the visual representation accurately reflected the actual displacement, without introducing any visual manipulation.

2) Procedure: Participants were asked to compare the stiffness of torsional springs rendered by three identical haptic interfaces, each coupled to the visual feedback displayed on the monitor. Participants provided feedback based on the perceived similarity to the reference. This task involved an ordinal classification and a pairwise comparison to evaluate the similarity of the stiffness renderings.

All optimization trials followed the same experimental procedure for each trial, as depicted in Fig. 5. In each trial, two different sets of parameters called the *query*, were sampled according to the UCB acquisition function (Query Generation in Fig. 5). Then, the participants were presented with the parameters of the query through the left and right haptic interfaces, while the reference model was rendered by the haptic interface in the middle (Execution in Fig. 5). Participants were asked to interact with the reference haptic interface at the beginning of each trial. After that, they were free to interact with the haptic interfaces on the left and right and go back to the reference, in any order they preferred. They were instructed to apply similar trajectories to all three knobs, ensuring they stay within the predetermined displacement range presented via the GUI as visual constraints. There was no time limitation for the trials.

Following the interactions with the haptic interfaces, participants were asked to compare the left and right knobs against the reference knob. Participants provided their responses to the two questions, based on their internal comparison (User Feedback in Fig. 5).

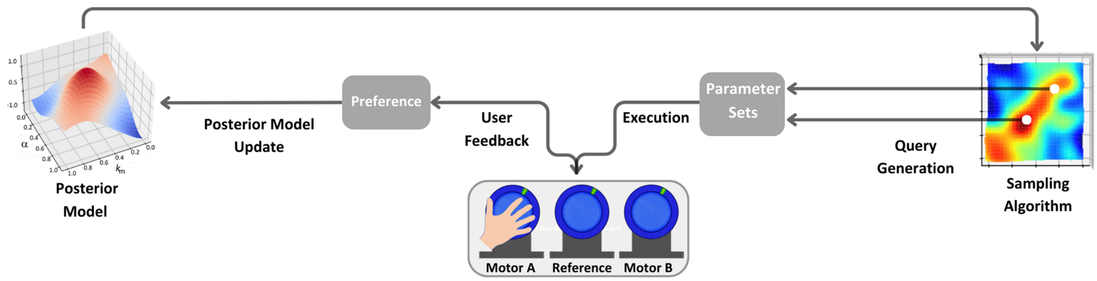

Fig. 5. A schematic representation of the preference-based HiL optimization.

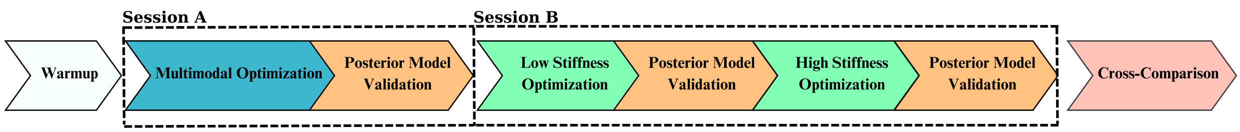

Fig. 6. A schematic representation of the experimental procedure.

For the first question, they evaluated the similarity of the feel of the right/left knobs, in comparison to the reference knob. For the ordinal classifications, they were asked to answer the question "*How do you feel when you interact with the knobs in comparison to the reference knob?*" by selecting one of the following responses:

- *Not Different:* the left (right) knob does not feel noticeably different compared to the reference knob.
- *Different:* the left (right) knob has a noticeably different feeling compared to the reference knob.
- *Very Different:* the left (right) knob has a significantly different feeling compared to the reference knob.

For the second question, they performed a pairwise comparison of the feelings of the left and right knobs with the reference knob and were asked to provide qualitative pairwise feedback by answering "*Which knob do you prefer?*" through one of the following responses:

- *Left:* the left knob feels more similar to the reference knob than the right knob.
- *Right:* the right knob feels more similar to the reference knob than the left knob.
- *Equal:* an equal level of similarity/dissimilarity exists for the left and right knobs compared to the reference knob.

Participants submitted their preferences through a GUI. Following this, the participants' preferences were utilized to update the GP for the perceived similarity with respect to the reference (Posterior Model Update in Fig. 5).

F. Sessions

The experiment was organized into three sessions: multimodal optimization, constant stiffness optimization, and cross-comparison sessions. The constant stiffness optimization session was designed to validate the findings acquired from the multimodal optimization.

Fig. 6 depicts the experimental flow: Each experiment started with a warm-up and ended with a cross-comparison. Ten-minute breaks were scheduled between the sessions and the volunteers were also allowed to take a break whenever they wanted. Additionally, the order of the multi-modal optimization (Session A) and the constant stiffness optimization (Session B) sessions, as well as the order of the low- and high-stiffness experiments, were randomized for each participant to prevent any patterns or learning effects during the experiments.

1) Warm-Up: The warm-up took about five minutes and was used to familiarize the volunteers with the haptic rendering task and the operation of the visual-haptic interface. During warm-up, all subjects were provided with at least six queries with 12 different renderings, while additional queries were provided until volunteers felt ready.

2) Constant Stiffness Optimization: The constant stiffness optimization sessions were designed to optimize the C/D ratio while keeping the stiffness parameter fixed at a predetermined value. The constant stiffness optimization consisted of two sub-sessions, denoted as high- and low-stiffness optimizations. These sub-sessions were configured to provide stiffness levels approximately 1.3 times higher and 1.3 times lower than the reference stiffness value of 0.2 Nm/rad, respectively.

The constant stiffness optimization sessions consisted of 12 trials for the low-stiffness and 12 trials for the high-stiffness experiments. It took a total of 36 iterations, including six trials for the low-stiffness posterior model validation and six trials for the high-stiffness posterior model validation. The constant stiffness optimization sessions with posterior model validations lasted about 30 minutes.

3) Multi-Modal Optimization: During the multi-modal optimization session, both the C/D ratio and the stiffness were concurrently optimized to explore the entire parameter space encompassing the C/D ratio and rendered stiffness variables.

The multi-modal optimization session consisted of 30 trials with 24 optimization trials and six posterior model validation trials. The multi-modal optimization session with the posterior model validation trials took about 30 minutes. The number of trials and the hyper-parameters used in the optimization were empirically decided based on pilot studies.

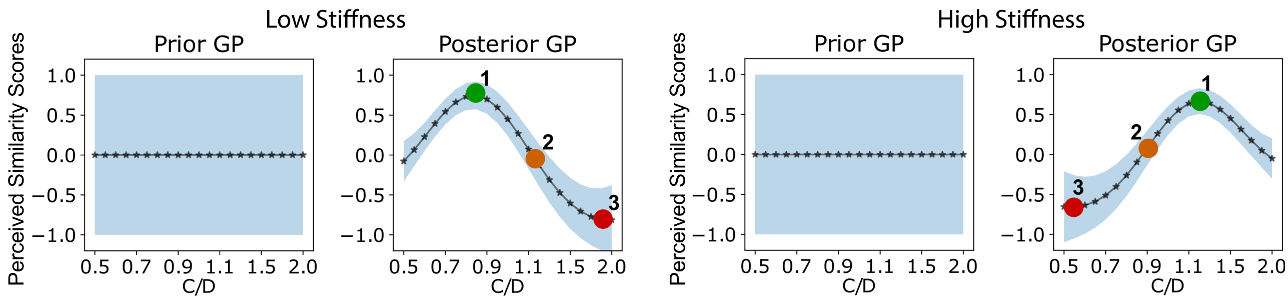

Fig. 7. Sample prior and posterior models (captured by their mean and standard deviation) for perceived similarity depicted for the one-dimensional experiments, where the stiffness was kept constant at 0.151 N-m/rad for low and 0.262 N-m/rad for high stiffness cases, and the C/D parameter was varied. The green, orange, and red points mark a local maximum, an intermediate point, and a local minimum of the posterior, representing equidistant perceived similarity scores in the search space including local maximum and local minimum levels.

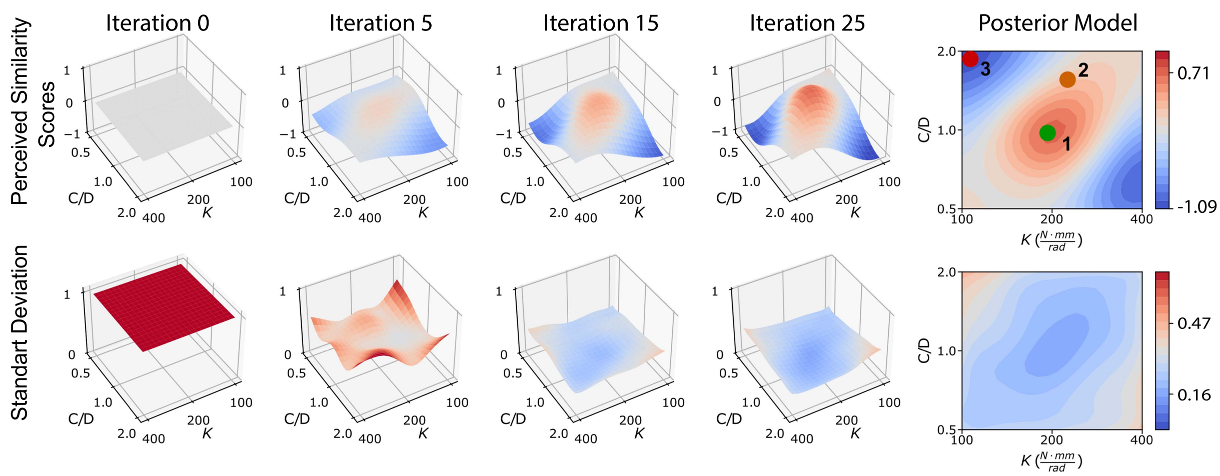

Fig. 8. The progression of the posterior model of multi-modal perceived similarity depicted at various trials of the HiL optimization for a participant. The first row captures the mean, while the second row presents the standard deviation of the posterior model. The green, orange, and red points mark the equidistant perceived similarity scores in the search space including the local maximum, intermediate, and local minimum values.

4) Posterior Model Validation: Posterior model validations were conducted after each constant stiffness optimization session and the multi-modal optimization session. Posterior model validations were performed by comparing non-deterministically selected test parameters that are representative of parameters resulting in a local minimum, a local maximum, and an intermediate perceived similarity score of the posterior model. The test parameters were rendered through the left or right knobs and compared to each other with respect to the reference knob.

Participants were asked to perform six pairwise comparisons of the local maximum, intermediate, and local minimum parameter sets for two blind orderings, for which the expected ordering was 1-local maximum, 2-intermediate, and 3-local minimum, as depicted in Figs. 7 and 8. These validation sessions were aimed to confirm the reliability of the posterior model derived through preference-based learning.

5) Cross-Comparison: After the completion of all optimization sessions, cross-comparison trials were conducted to test if participants could distinguish their optimized parameters from the reference parameter. During the cross-comparison, participants were asked to compare two rendering models with each other presented by the left and right knobs with respect to their similarity to the reference at the middle knob.

Participants followed the same procedure with the optimization sessions to provide pairwise and ordinal feedback as in Section IV-E2. The queries were not generated by the acquisition function, instead, the optimized parameter sets for the low-, high-stiffness, and multi-modal settings, together with the reference parameters, were utilized. Each comparison was repeated two times for each parameter set, and all comparisons were completed in 12 trials, resulting in six pairwise comparisons and six ordinal classifications for a participant. The cross-comparison session lasted about 10 minutes.

G. Data Analysis

To check for statistically significant differences in the ordinal classifications of the low-stiffness, high-stiffness, and multi-modal optimized parameter sets tested during the cross-comparison sessions, first, the classification outcomes of participants were recategorized into a binary variable, in which either the participant perceived the rendering as *similar* to the reference or *different* from the reference, with the latter combining the *different* and *very different* categories. Then, a logit model was utilized for the analysis as the dependent variable was derived from the binary classification outcomes.

The explanatory variables of the logit model consist of the type of optimal parameter sets (low-stiffness optimal, high-stiffness optimal, multi-modal optimal) and an additional handedness variable introduced to capture any systematic difference in the propensity to consider the left knob as similar to the reference, relative to the right knob. The handedness variable ensures the model accounts for potential asymmetries in participants' perception between the left and right knobs.

The logit model captures the probability of a participant perceiving the rendering as similar to the reference as a function of the rendering parameter sets and the knob position. The logit model also includes an intercept, which captures the average effect of all variables not explicitly included in the model that contribute to the propensity rate of the rendering to be categorized as similar to the reference. Given the setup of the model, the intercept effectively corresponds to the probability of selecting *similar* when the rendering parameter set and the knob position are at their reference values.

To account for the dependency within each participant's observations due to the repeated-measures nature of the experiment design, cluster-robust standard errors were used. Cluster-robust errors adjust for the intra-cluster correlation by treating each participant as a cluster, ensuring that the standard errors properly reflect the variability across the participants rather than within individuals, providing more reliable inferences.

For the posterior validation, the performance of all three learned perception models was studied by confusion matrices. The parameters corresponding to the local maximum, local minimum, and intermediate values of the learned GP posterior model, along with the user ordering, were utilized for multi-class classification.

V. Results

A. HiL Optimizations

The GP prior and posterior models, captured by their mean and standard deviation, for a sample participant during the low- and high-stiffness rendering experiments are presented in Fig. 7. The progressions of the GP over the trials during multi-modal experiments for a sample participant are presented in Fig. 8. The last columns of Figs. 7(a) and (b) and 8 also present the orderings of the participants, corresponding to the local maximum (1), intermediate (2), and local minimum (3) of the posterior.

The average optimized C/D ratio of the 12 volunteers for the low-stiffness setting is determined as 0.77 ± 0.12, while the average C/D value for the high-stiffness setting is identified as 1.35 ± 0.18. The average stiffness and C/D ratio values of the multi-modal setting are determined as 0.20 ± 0.03 (Nm/rad) and 1.06 ± 0.25, respectively.

B. Posterior Model Validation

To validate the optimized perception models, the participants ordered the representative local minimum, intermediate, and local maximum values of the posterior models. GP orderings were considered the ground truth and were used along with blind orderings of the participants for the multi-class classification. The accuracy of the orderings is presented as confusion matrices in Fig. 9. Each cell in the matrices indicates the normalized number of correct classifications. Since the participants were allowed to indicate ties, the rows/columns of the confusion matrices may not sum to one. The dominance of the diagonals shows how well the Gaussian posterior model can capture the underlying perception model.

TABLE I
Results of the Statistical Analysis for Cross-Comparisons

	estimate	std. error	t-ratio	statistically significant
Intercept	1.332	0.403	3.308	yes – at 1% level
Low-stiffness	-1.766	0.553	-3.191	yes – at 1% level
High-stiffness	-1.310	0.546	-2.400	yes – at 5% level
Multi-modal	-0.743	0.753	-0.986	no
Handedness	-0.156	0.284	-0.549	no

C. Cross-Comparison

As a result of the cross-comparison sessions, each optimal parameter set was classified 72 times. The optimal parameter sets were labeled as similar to the reference parameter set for 43.1%, 48.6%, and 68.1% of these trials for the low-stiffness, high-stiffness, and multi-modal cases, respectively.

Table I presents the parameter estimates according to the logit model detailed in Section IV-G, with their corresponding cluster-robust standard errors and the t-statistics, calculated as the parameter estimates divided by their respective standard errors. The last column denotes the statistical significance together with the corresponding confidence levels.

The results for the *intercept* show that the baseline probability of a participant selecting a knob with the reference parameters being similar to the reference rendering is statistically significant at the 1% confidence level, with an odds ratio of approximately 3.8 to 1.

The results for the *low-stiffness* optimal parameter sets indicate a statistical significance at the 1% confidence level. Hence, there exists strong evidence indicating that the null hypothesis—the low stiffness and the reference stiffness renderings are perceived as similar—can be rejected.

Similarly, the results for the *high-stiffness* optimal parameter sets indicate a statistical significance at the 5% confidence level. Hence, there exists sufficient evidence indicating the null hypothesis—the high stiffness and the reference stiffness renderings are perceived as similar—can be rejected.

The results for the *multi-modal* optimized parameter sets do not indicate any statistical significance. Hence, the null hypothesis—-the multi-modal stiffness and reference stiffness rendering are perceived as similar—-cannot be rejected.

Finally, the results for the *handedness* do not indicate any statistical significance; hence, the null hypothesis of their equivalence cannot be rejected.

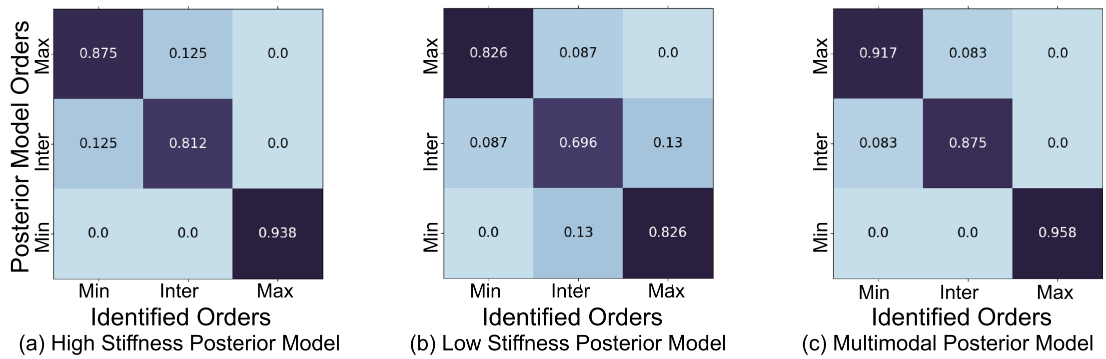

Fig. 9. Confusion matrices of the posterior model validation.

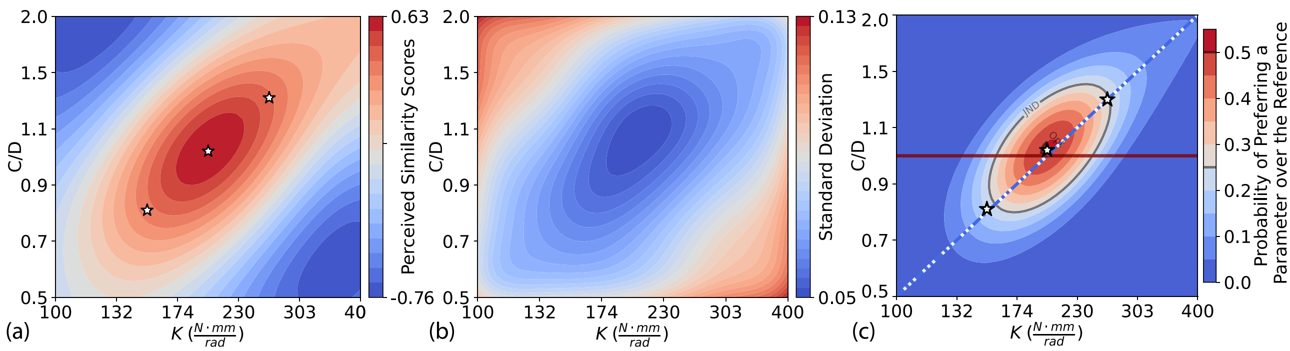

Fig. 10. (a) The mean values of the averaged GP posterior, for which the red regions represent higher perceived similarity scores. The stars depict the mean parameter values for the low stiffness, high stiffness, and multi-modal optimizations. (b) The standard deviation of the averaged GP posterior, for which higher values are attained in the red regions, where the algorithm samples less frequently. The red regions are sampled less frequently since the algorithm estimates these regions to have lower perceived similarity scores than the blue regions in the standard deviation plot. (c) The probability map of selecting a parameter set in pairwise comparison with respect to the reference. The region inside the JND curve captures the parameters that can be used instead of the reference parameters. The lines on the figure are used to analyze the effect of the C/D ratio over the perceived stiffness.

D. Averaged Gaussian Process Posterior

The averaged multi-modal perception model, computed according to Section III-4a, which provides the underlying perceived similarity ratings for 12 participants is presented in Fig. 10. In particular, Fig. 10(a) presents the mean value plot annotated with the average values of the high, low, and multi-modal stiffness optimization results, while Fig. 10(b) presents the standard deviation of the averaged GP posterior.

E. Just Noticeable Difference Thresholds

The estimated JND thresholds for perceived stiffness are depicted in Fig. 10(c). Fig. 10(c) presents a re-interpretation of Fig. 10(a) and (b) through the probability mapping discussed in Section III-4b. The contours in this figure capture the probability of the rendered parameters being indistinguishable from the reference, as well as the boundary of JND.[1] The average values for the low, high, and multi-modal stiffness optimization results are also annotated with stars in the figure.

[1]Probabilities exceeding 50% in Fig. 10(c) are artifacts of the slight mismatch between the multi-modal optimized and the reference stiffness.

VI. Discussion

In this section, we elaborate on each hypothesis presented in Section IV-D and compare our results with the related studies in the literature.

A. Hypothesis 1

The preference-based learning algorithm simultaneously constructs a Gaussian latent model of the perceived similarity of each individual, as depicted in Figs. 7 and 8. Accordingly, it becomes possible to assign a perception score to any parameter set within the design space employing Gaussian regression, even if that particular parameter set has not been directly assessed.

The posterior model validation results indicate that the GP model successfully captures the perceived similarity of the participants, as evidenced by the diagonal dominance in the confusion matrices in Fig. 9. These results demonstrate the efficiency of the GP in modeling the latent perception model, while concurrently optimizing the rendering model parameters.

Low standard deviations in the posterior plots indicate adequate convergence of the GP models. Compared with the standard deviation of the averaged model prior, the standard deviation of the averaged model posterior has reduced by more than 80% in the blue regions, indicating a high information gain. Accordingly, the blue dominant region Fig. 10(b) indicates where the averaged multi-modal perception model can be utilized with high confidence.

The red dominant region of the averaged posterior GP model in Fig. 10(a) indicates the stiffness and C/D ratio parameters for which the perceived similarity of the rendering compared to the reference is evaluated to be high. The average posterior model shows that the parameter sets close to the counter-diagonal (from the bottom left corner to the top right corner) defining the visual stiffness K_v have the highest perceived similarity scores, compared to the other parameter sets.

Overall, the results provide strong evidence that proportionally increasing (decreasing) the C/D ratio can cause a haptic interface to feel more compliant (stiffer). Accordingly, the results support the first hypothesis H1.

B. Hypothesis 2

While the averaged posterior GP model in Fig. 10(a) provides evidence of the feasibility of perceived stiffness modulation via visual scaling, further insights may be gained about the quality of such renderings by inspecting the perceived similarity scores compared to the reference model.

The averaged posterior model in Fig. 10(a) captures a significant decrease in the perceived similarity scores as the C/D ratio deviates further from the unity. For instance, the cross-comparison results of the multi-modal, high-stiffness, and low-stiffness optimization indicate that on average these parameter sets were declared to be similar to the reference model for 68.1%, 48.6% and 43.1% of the trials, respectively. Note that these results should be compared with the odds of the intercept, indicating that the reference parameters were declared to be similar to the reference for 79% of the trials.

Fig. 10(c) presents the JND threshold estimated for the perceived stiffness according to Section V-E. The area inside the JND threshold ellipsoid captures the parameter sets that are indistinguishable from the reference rendering, while the parameter sets outside this closed curve have detectable differences. Please note that since users were not asked to simply compare stiffness but to rate the overall perceived similarity of the rendering, the existence of a detectable difference does not necessarily imply that the perceived stiffness of the rendered set is evaluated to be different; it only implies that users can detect some difference between the renderings, for instance, possibly due to the detection of manipulated visual cues.

The JND threshold for the perceived stiffness depicted in Fig. 10(c) provides an explanation for the cross-comparison results. According to Fig. 10(c), the average of the multi-modal optimal parameter sets falls well inside the JND region, while the average of the high-stiffness optimization results resides on the JND boundary, and the average of the low-stiffness optimization results falls outside of the JND boundary. Hence, on average, the multi-modal optimization parameter sets are not expected to be distinguishable from the reference model, the high-stiffness optimization parameter sets are expected to be distinguishable from the reference model at about the chance level, while the low-stiffness optimization parameter sets are expected to be easily distinguishable.

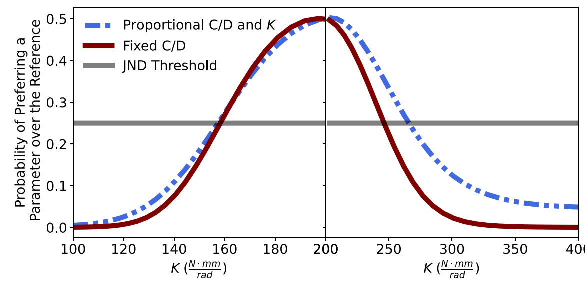

Fig. 11. Probability of parameters to be preferred over the reference, computed by considering slices of Fig. 10(c) along the straight lines depicted in that figure. The horizontal grey line depicts the JND threshold.

In line with these observations, the cross-comparison results indicate that the multi-modal optimal parameter sets are not statistically significant, supporting the high perceived similarity with respect to the reference. Furthermore, the statistical tests for high-stiffness and low-stiffness optimal parameter sets indicate statistically significant differences with respect to the reference, at 1% and 5% confidence levels, respectively. Given that the high-stiffness optimal parameter sets lie on the JND threshold while low-stiffness optimal parameter sets are outside the JND threshold, both the detection of the significant differences for both parameter sets and an increase in the statistical confidence levels from high- to low-stiffness optimal parameter sets are consistent with our model.

We further study the effect of visual-haptic incongruency on the JND thresholds by considering slices of Fig. 10(c) along the straight lines depicted in the figure. The solid red line in Fig. 10(c) captures the case when the C/D ratio is set to unity, while the dashed blue line considers the case when the stiffness and C/D ratio parameters are increased proportionally such that the visual stiffness is kept constant at $K_v = K_{ref}$. Fig. 11 presents the cross-sections of these slices characterizing the effect of visual cue modulation on stiffness perception. The horizontal gray line in Fig. 11 depicts the JND threshold.

The virtual stiffness JND level in Fig. 11 for the unity C/D ratio, corresponding to changing stiffness without any visual manipulation, is consistent with the virtual stiffness perception thresholds reported in the literature [54], [55]. The blue dashed curve in Fig. 11 represents a proportional relationship between stiffness and C/D ratio parameters and provides supporting evidence that the perception of stiffness can be enhanced through visual scaling with high perceived realism. The blue dashed curve indicates that as the C/D ratio further deviates from unity, visual incongruency increases and finally interferes with the perceived realism of rendering. In particular, while increasing the perceived compliance, a proportional increase of the C/D ratio can extend the JND range from 20% to 35% of the reference stiffness in comparison to the case when the C/D ratio is set to unity. Hence, the JND threshold can be significantly increased through visual-haptic incongruency. Accordingly, by increasing the C/D ratio, one can cause a stiff haptic interface to feel more compliant in a manner that is indistinguishable from the reference model.

However, the range of stiffness modulation with high perceived realism is not symmetrical; a similar effect through visual

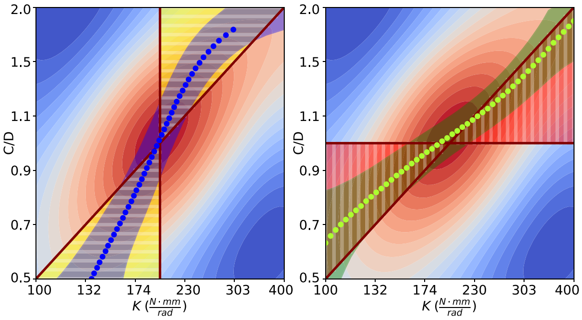

Fig. 12. (a) The yellow area indicates the region of sensory integration for constant C/D ratios with variable stiffness. Blue points have the highest perceived similarity score along the horizontal axes, with the blue-shaded region around these points denoting one standard deviation confidence interval. (b) The red area indicates the region where sensory integration for constant stiffness with variable C/D ratios. Green points have the highest perceived similarity score along the vertical axes, with the green-shaded region around these points denoting one standard deviation confidence interval.

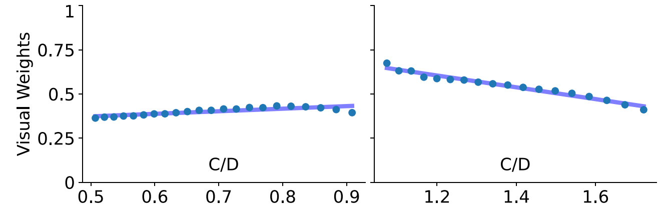

Fig. 13. Estimation of the weight w_v of visual cues utilized during sensory integration with respect to C/D ratio.

cue manipulation (decreasing the C/D ratio) is not observed (with a meaningful effect size) when aiming to achieve stiffer perceived renderings that are indistinguishable from the reference stiffness. In particular, under visual attenuation, while the users perceive the stiffness to be higher than the haptic stiffness, they also indicate that they can detect a difference compared to the reference model. This detectable difference is possibly due to the experimental setup necessitating excitations from the forearm rotations. A low C/D ratio requires a larger range of motion to achieve a similar visual excitation and this causes the forearm to approach its joint limits. Furthermore, such a scaling requires humans to perform more work for the same visual input. Overall, these effects may result in a detectable difference compared to the reference rendering when a C/D ratio is set lower than unity.

In summary, our results suggest that while high C/D ratios can increase perceived compliance in a manner that is indistinguishable from the reference model, low C/D ratios can increase the perceived stiffness, but users evaluate the overall perceived realism to be distinguishable from the reference model. Hence, the perceived realism due to visual-haptic incongruency is not symmetrical and favors increasing the perceived compliance.

Overall, the cross-comparison results and the JND threshold in Figs. 10(c) and 11 demonstrate that the perceived realism under visual-haptic incongruency can be kept high for a limited range of visual scaling. Accordingly, the results support the second hypothesis H2.

C. Hypothesis 3

In Fig. 12, to study sensory integration using the averaged posterior model, we consider vertical slices along Fig. 10(a) to determine the best C/D ratio for a given fixed stiffness and horizontal slices along Fig. 10(a) to determine the best stiffness for a given C/D ratio, respectively.

Along the horizontal lines, as depicted in Fig. 12(a), where a constant C/D ratio is considered, users determine the best stiffness that reduces the perceived difference between the rendered parameters and the reference rendering. If the users neglect the scaled visuals and only rely on the movement-related cues from the haptic modality, then the perceived stiffness will be the haptic stiffness K_h. If the users utilize only the visual movement cues presented on the screen and neglect the movement-related cues from the haptic modality, then the perceived stiffness will be the visual stiffness K_v and lie along the counter-diagonal in Fig. 12(a). If users utilize movement-related cues from both haptic and visual modalities and if sensory integration is employed, then the perceived stiffness will lie in the yellow triangular areas covering the range between the haptic stiffness K_h and the visual stiffness K_v.

Similarly, along the vertical lines, as depicted in Fig. 12(b) where a constant stiffness is considered, users determine the best C/D ratio that reduces the perceived difference between the rendered parameters and the reference rendering. If users select a C/D ratio only to reduce the difference between haptic and visual movement, then the ideal C/D ratios will lie along the horizontal line. If users are not affected by the movement difference between haptic and visual modalities and select C/D ratios to match the visual stiffness to the reference, then the ideal C/D ratios will lie along the counter-diagonal in Fig. 12(b). If sensory integration is employed, then the C/D ratio will lie in the red triangular areas covering the range between the horizontal and the counter-diagonal.

The points marked in Fig. 12(a) and (b) depict the points with the highest probability of similarity to the reference along their respective (horizontal and vertical) search directions, while the shaded regions around these points denote one standard deviation confidence intervals. Since these points captured by our probability map reside inside the colored triangular areas, there exists strong evidence for sensory integration between visual and haptic cues during stiffness perception.

Furthermore, the points in Fig. 12 can also be used to estimate the relative contributions of haptic and visual cues utilized by stiffness perception under visual-haptic incongruency, if a weighted linear combination of the cues is assumed for (possibly partial) integration. For instance, Fig. 12(a) captures the case of sensory integration between the haptic stiffness $K_h = K$ and the visual stiffness $K_v = \frac{K}{\text{C/D ratio}}$. Given the points in Fig. 12(a) satisfy the subjective equality, called PSE [5], the visual weights can be estimated according to the linear interpolation $w_v = (PSE - K_h)/(K_v - K_h)$, while the haptic weights can be computed as $w_h = 1 - w_v$.

Fig. 13 presents the estimates of the contribution (or weight) w_v of visual cues utilized during sensory integration with respect to the C/D ratio, computed according to Fig. 12(a). Only the parameters corresponding to the adequately explored regions of the parameter space are included in Fig. 13 to preserve the reliability of the estimates.[2]

The results indicate that as the C/D ratio deviates further from unity, increasing the level of incongruency, the contribution w_v of visual cues in sensory integration decreases. The crossing of the 50% threshold in Fig. 13 further indicates that the dominant sensory modality for the movement-related cues switches from vision to haptics as the C/D ratio deviates from the unity. Moreover, the visual contributions are relatively higher for the range of C/D ratios aimed to increase perceived compliance, when compared to the range of C/D ratios aimed to increase perceived stiffness.

These weights explain the non-symmetrical nature of the perceived realism presented in Fig. 11, indicating that high C/D ratios result in a more rapid decrease in the contribution of the visual cue. Finally, an extrapolation of the decreasing contribution trend in Fig. 13 suggests that vetoing of visual modality is likely to occur for large incongruency levels. These results strongly support the last hypothesis H3.

D. Comparisons With the Related Work

Our results are in good agreement with [1], [5], [15], [16], indicating that the sensory integration process during multi-modal visual-haptic compliance estimation under visual scaling can be modeled through a weighted linear combination of redundant sensory signals and manipulated visual cues can expand the perceived stiffness range of haptic rendering.

While our study is similar to [15], [16], we explore a continuous range of visual scaling levels from a perceived similarity perspective and our results are in more line with those of [1], [5]. Our findings extend [16] and support [1], [5], indicating a decreased contribution of the manipulated visual cue in the multi-modal percept as its congruency is reduced, and suggests possible vetoing of this unreliable cue when strong conflicts arise between the haptic modality and the manipulated visual feedback.

Our study supports the conclusions of [13], [16], [19], indicating that pseudo-haptics can enhance haptic feedback during stiffness rendering. Our approach differs as we rely on qualitative feedback-based HiL optimization to maximize the perceived similarity with respect to the reference and determine the optimal levels of the visual scaling.

Our results support the observations in [20], [22] stating visual-haptic incongruency significantly influences stiffness discrimination performance when the discrepancy between the visual feedback and the hand movements is large. We extend these results by explicitly characterizing the relationship between the incongruency levels and the perceived realism.

[2]C/D ratios near the boundaries of our search space are omitted from the estimates as these parts of the probability map are under-explored. Similarly, the C/D ratio near unity is not included as the estimation becomes very sensitive to experimental noise around this value.

Our JND estimations are consistent with the results of earlier psychophysical experiments, which reported the JND threshold as 23% [55] or in the $8\% - 22\%$ range [54].

For the comparison of our results with the discrimination threshold models in [56], Fig. 10(c) can be mapped to a visual stiffness versus haptic stiffness increment (ΔK_h–ΔK_v) plot, where both stiffness increments are defined with respect to K_{ref}. After this transformation, the JND ellipsoid along the counter-diagonal of Fig. 10(c) maps to another ellipsoid along the main diagonal (from the top right corner to the bottom left corner) of the new plot, since K_v is inversely proportional to the C/D ratio. The mapped JND ellipsoid in the ΔK_h–ΔK_v plot is consistent with the discrimination threshold model proposed in [56] when participants have access to both the single-cue and combined estimates.

Furthermore, given the parameters with the highest perceived realism scores inside the JND threshold (residing predominantly along the counter-diagonal) in Fig. 10(c) are indistinguishable from the reference, they can be classified as metamers—physically distinct, but perceptually indiscriminable, stimuli [56], [57].

While our study determines the relative contributions of visual and haptic modalities with respect to the visual scaling under the assumption of a weighted linear combination of these cues, it makes no attempt to determine the underlying model of (possibly partial) sensory integration.

One interpretation of the results in Fig. 13 is possible by utilizing the causal inference model [12], which generalizes the MLE model of multisensory cue integration by considering the causes of the underlying sensory events. In particular, the causal inference model infers the probability of a common cause versus two independent causes to derive optimal predictions. Accordingly, this model involves two parameters, capturing the probability of a common cause and the relative weight of each cue for the sensory integration case.

On the one hand, the causal inference model reduces to maximum likelihood estimation when there is a high probability of a common cause. In our experiments, participants were provided explicit instructions emphasizing the common cause between the visual and haptic modalities. Furthermore, while the visual feedback was scaled, the synchronization between the haptic input and visual output was preserved throughout the experiments, resulting in a convincing causal reason for a common cause. Accordingly, if one considers that the probability of independent cause was low and the discrepancies were sufficiently small and hard to detect, then the weights in Fig. 13 may be attributed to the weights of an MLE-type sensory integration.

On the other hand, due to the experiment design, our results do not allow for the extraction of the two parameters of the causal inference model from the data; hence, the common cause assumption cannot be verified. Furthermore, discrepancies may become large as the C/D ratio further deviates from unity. As an alternative interpretation, if one assumes that the visual feedback completely dominates the movement-related cues, then the weights in Fig. 13 can be interpreted as the probability of a common cause.

Overall, it is likely that the underlying model of sensory integration lies somewhere between these two interpretations, performing partial sensory integration under conflicting cues.

VII. Conclusion and Future Work

We provide evidence that the perception of users can be successfully manipulated by changing the visual modality without altering the rendered stiffness parameter, to make virtual environments feel stiffer or more compliant. We show that HiL optimization based on qualitative feedback is an effective means of customizing the rendering parameters under visual-haptic incongruency to achieve consistently high perceived realism scores for a given haptic interface and a user.

We demonstrate that the level of visual-haptic incongruency introduces limitations to the range of visual scaling for which the perceived realism can be kept high. We provide evidence that while increasing compliance via amplification of the visual feedback has high perceived realism, increasing stiffness via attenuation of the visual feedback provides lower perceived realism, as effects such as the joint limits of the user and extra work that needs to be performed to achieve a similar visual excitation may result in detectable differences. To determine these limits of performance, we provide a novel means of estimating the JND thresholds through the HiL optimization.

Finally, we provide strong evidence that movement-related cues from the haptic and visual modalities are integrated by a weighted linear combination to form a stiffness percept from the redundant sensory modalities. Accordingly, the perceived stiffness is formed as a linear combination of ideal visual stiffness and haptic stiffness values. We provide a novel means of estimating these weights in visual-haptic sensory integration and show that as the incongruency level of the visual feedback increases, the contribution of visual cues (their respective weights in the stiffness perception) decreases.

Overall, we advocate for the utilization of HiL optimization during multi-modal haptic rendering, since it not only allows for the determination of appropriate haptic-visual parameters to ensure a consistently high rating of perceived realism but also provides an efficient means to study multi-modal sensory integration under conflicting cues.

A. Study Limitations and Future Work

While the proposed non-parametric Bayesian optimization framework provides insight into the sensory integration process under conflicting cues and the results that are in good agreement with the related works, there are many interesting research directions along which the proposed method can be further improved.

The main goal of this study was determined as HiL optimization of the scaling ratio to maximize the perceived realism of stiffness rendering under visual-haptic incongruency. Accordingly, the acquisition function has been selected in such a way that it prioritizes the sample efficiency, favoring exploitation over exploration. Hence, while the HiL optimization adequately samples the regions near the maxima, non-promising regions may be under-sampled. This choice results in fast convergence of the approach to the optimal parameters.

As a trade-off of this choice, the probability estimates in the less explored region of the parameter may become less reliable. To extend the range of adequately explored parameter space, one can utilize the HiL optimization approach with an alternative acquisition function that provides better exploration properties or include uniform random sampling for the unexplored regions of the parameter space, at the cost of a longer experiment duration.

Similarly, the ordinal and pairwise feedback collected from the user in our study were designed to maximize the perceived similarity of stiffness rendering with respect to the reference model. Accordingly, while our JND estimations are consistent with the results of psychophysical experiments reported in the literature [54], [55], [56], these questions do not necessarily constitute the most direct queries for determining JND thresholds. Queries may be updated if the determination of the JND thresholds is the main goal.

During the non-parametric modeling with GPs, the selection of the underlying kernel and its hyperparameters, as well as the enforcement of certain constraints over the latent function, such as monotonicity, may improve the optimization process. For simplicity and flexibility in analysis, we utilized a GP with a radial basis kernel with no constraints, and the resulting latent function was captured as monotonic. Future studies may impose monotonicity, possibly along a single stimulus dimension, to further improve the sample efficiency.

Furthermore, in this study, while the kernel parameter was optimized to achieve the maximum likelihood of observations, the rest of the hyperparameters of the GP model were selected based on insights gained from preliminary experiments. The utilization of more advanced hyperparameter tuning methods is a part of our future work.

The anonymized data collected during our study, the HiL optimization algorithm, and the values of empirically determined hyperparameters are available at GitHub[3] to enable other researchers to build upon our study.

Appendix A
Active Learning Model and Bayesian Optimization

1) Latent Function Modeling: Let $\boldsymbol{X} \subset \mathbb{R}^{d\times d}$ be the finite set of parameters used in visual-haptic rendering. Let $\boldsymbol{x} = \{x_i : i = 1, \ldots, 2n\}$ and $\boldsymbol{x} \subset \boldsymbol{X}$ be the rendered parameter sets observed by the subject in the HiL trials until the n^{th} iteration. Let $f(\boldsymbol{x})$ be the latent function that reflects the human perception for the rendered parameters $\boldsymbol{x}$. Since GP regression can closely approximate black-box functions, such as psychometric field $f(\boldsymbol{x})$, a normalized GP is modeled as

$$f(\boldsymbol{x}) \sim GP\left(f(\boldsymbol{x}); 0, \Sigma(\boldsymbol{x})\right) \quad (5)$$

$$P(f(\boldsymbol{x})) = \frac{1}{(2\pi)^n |\Sigma(\boldsymbol{x}))|^{\frac{1}{2}}} e^{-\frac{1}{2}(f(\boldsymbol{x})\Sigma(\boldsymbol{x})^{-1} f(\boldsymbol{x}))} \quad (6)$$

[3]https://github.com/HarunTolasa/Gaussian-Process-Library

where $\Sigma(\boldsymbol{x})$ represents the covariance matrix

$$\Sigma(\boldsymbol{x}) = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \cdots & k(x_n,x_n) \end{bmatrix} \tag{7}$$

and the covariance between each parameter set is defined by a radial basis kernel defined as

$$k(x_i,x_j) = e^{-\gamma\|x_i - x_j\|^2}. \tag{8}$$

During a trial, subjects provide two separate ordinal classifications for two rendering parameter sets and one pairwise preference between them. Let q_i be a subject's feedback for i^{th} iteration; $q_i \subset \{q_{o_{i_1}}, q_{o_{i_2}}, q_{p_i}\}$. Let $D \subset \{q_i : i = 1,\ldots,n\}$ be the total set of qualitative feedback data.

If the latent function is known, then the probabilities of subjects providing ordinal classification and pairwise preference correctly are $P(q_o|f(\boldsymbol{x}))$ and $P(q_p|f(\boldsymbol{x}))$, respectively. Therefore, for a known latent function, their ability to provide the correct answer for the i^{th} iteration has the probability

$$P(q_i|f(\boldsymbol{x})) = P(q_{oi_1}|f(\boldsymbol{x}))P(q_{oi_2}|f(\boldsymbol{x}))P(q_{pi}|f(\boldsymbol{x})). \tag{9}$$

Assuming that all of the qualitative feedback is independent of each other, then the probability of the collected qualitative feedback data up to n^{th} iteration being correct in HiL optimization is given by

$$P(D|f(\boldsymbol{x})) = \prod_{i=1}^{n} P(q_i|f(x)) \tag{10}$$

Let ϵ_p in Fig. 2 be a Gaussian white noise with variance c_p^2. Then, the probability of participants providing correct pairwise comparison is defined as

$$P(q_{pi} = (x_{i1}) \succ (x_{i2})|f(x)) = \Phi\left(\frac{f(x_{i1}) - f(x_{i2})}{c_p}\right) \tag{11}$$

where Φ represents the cumulative distribution function of the Gaussian distribution.

Let $O = \{o_1, o_2 \ldots, o_r\}$ be the finite set of r ordinal classifications and t_{o_j} be the threshold of each ordinal classification $-\infty = t_{o_0} < t_{o_1} < t_{o_2} < \ldots < t_{o_r} = \infty$. Let ϵ_o in Fig. 2 be a Gaussian white noise with variance c_o^2. Then, the probability for parameter x_i belonging to the o_j^{th} ordinal class is defined as

$$P(q_{oi} = o_j|f(x_i)) = \Phi\left(\frac{t_{o_j} - f(x_i))}{c_o}\right) - \Phi\left(\frac{t_{oj-1} - f(x_i)}{c_o}\right) \tag{12}$$

where c_o is the noise level constant.

2) Posterior Latent Function and Bayesian Inference: The posterior probability distribution function of $f(\boldsymbol{x})$ for the collected qualitative feedback data D be modeled as

$$P(f(\boldsymbol{x})|D) \propto P(D|f(\boldsymbol{x}))P(f(\boldsymbol{x})). \tag{13}$$

The Laplace method is used to approximate the posterior distribution of $P(f(\boldsymbol{x})|D)$ as a multi-variate Gaussian distribution. The Laplace approximation uses a second-order Taylor expansion of posterior distribution around its mode [44]

$$f(\boldsymbol{x})_{|D} \sim GP(f(\boldsymbol{x}); \hat{f}, (W + \Sigma(\boldsymbol{x})^{-1})^{-1}) \tag{14}$$

$$f(\hat{\boldsymbol{x}}) = \text{argmax}_{f(\boldsymbol{x})}\left(log(P(D|f(\boldsymbol{x}))P(f(\boldsymbol{x})))\right) \tag{15}$$

where W is the negative Hessian of $log(P(q|f(\boldsymbol{x})))$ defined as

$$W_{ij} = -\frac{\partial^2 \, log(P(D|f(\boldsymbol{x}))}{\partial f(x_i)\, \partial f(x_j)}.$$

From the posterior distribution of $f(\boldsymbol{x})|q = f_{|D}$, the output of the latent function $f(x_*) = f_*$ for any arbitrary parameter set $x_* \subset \boldsymbol{X}$ can be estimated. Let $f_{*|D}$ be the estimated output of the latent function for an arbitrary parameter set x_* based on the information q. Then, $f_{*|D}$ is another Gaussian distribution that can be computed by marginalization.

$$P(f(x_*)|D) = \int_{-\infty}^{\infty} \left(P(f(x_*)|f(\boldsymbol{x}), q)P(f(x)|D)\right) df(\boldsymbol{x}) \tag{16}$$

The result of the integral in (16) is given as

$$f_{*|D} \sim GP(f; E(f_{*|D}), Var(f_{*|D})) \tag{17}$$

with

$$E(f_{*|D}) = k_*\,\Sigma^{-1}\hat{f}$$
$$Var(f_{*|D}) = k_{**} - k_*\,(\Sigma + W^{-1})^{-1}k_*^T.$$

where $k_* = k(x_*, x_{1:n})$ and $k_{**} = k(x_*, x_*)$. The estimated mean and variance of $f_{*|D}$ in (17) are utilized by the acquisition function to predict the most promising points for the next iteration, as well as to form an understanding of the perceived similarity of the participants.

3) Predicting Subject Preferences:: Marginalization also allows for the prediction of posterior probabilities of qualitative feedback. We used it to find the posterior probability of preferring an arbitrary parameter x_* over the reference x_{ref}, denoted with probability $P(x_* \succ x_{ref}|D)$. Let Δf be the difference between latent scores $f_* - f_{ref}$. For known a known difference Δf, the preference probability of the user is modeled as

$$p(x_* \succ x_{ref}|\Delta f) = \Phi(\Delta f/c_p) \tag{18}$$

The posterior distribution of latent score difference Δf has a Gaussian distribution with the mean $E(\Delta f_{|D})$ and the variance $Var(\Delta f_{|D})$.

$$E(\Delta f_{|D}) = E(f_{*|D}) - E(f_{ref|D}) \tag{19}$$

$$Var(\Delta f_{|D}) = Var(f_{*|D}) + Var(f_{ref|D}) - 2Cov(f_{*|D}, f_{ref|D}) \tag{20}$$

The mean and variance of $\Delta f_{|D}$ is found using (17). Then, $P(x_* \succ x_{ref}|D)$ is estimated as

$$P(x_* \succ x_{ref}|D) = \int_{-\infty}^{\infty} \Phi\left(\frac{(\Delta f)}{c_p}\right) P(\Delta f|D) d\Delta f \tag{21}$$

and the result of this integration is denoted as

$$P(x_* \succ x_{ref}|D) = \Phi\left(\frac{E(\Delta f_{|D})}{\sqrt{Var(\Delta f_{|D}) + c_p^2}}\right). \tag{22}$$

ACKNOWLEDGMENT

The authors thank Seda Erdem and Danny Campbell for their valuable assistance in the statistical analysis of the cross-comparison results and Cagatay Basdogan for bringing some related studies on sensory cue integration to our attention. The authors also thank anonymous reviewers for their comprehensive constructive feedback that helped us significantly improve our manuscript.

REFERENCES

[1] M. O. Ernst and H. H. Bülthoff, "Merging the senses into a robust percept," *Trends Cogn. Sci.*, vol. 8, no. 4, pp. 162–169, 2004.

[2] R. v. Ee, W. J. Adams, and P. Mamassian, "Bayesian modeling of cue interaction: Bistability in stereoscopic slant perception," *J. Opt. Soc. Amer. A*, vol. 20, no. 7, pp. 1398–1406, 2003.

[3] J.-P. Bresciani, F. Dammeier, and M. O. Ernst, "Vision and touch are automatically integrated for the perception of sequences of events," *J. Vis.*, vol. 6, no. 5, pp. 554–564, 2006, doi: 10.1167/6.5.2.

[4] M. J. Young, M. S. Landy, and L. T. Maloney, "A perturbation analysis of depth perception from combinations of texture and motion cues," *Vis. Res.*, vol. 33, no. 18, pp. 2685–2696, 1993.

[5] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, pp. 429–433, 2002.

[6] S. Gepshtein and M. S. Banks, "Viewing geometry determines how vision and haptics combine in size perception," *Curr. Biol.*, vol. 13, no. 6, pp. 483–488, 2003.

[7] H. H. Bülthoff and H. A. Mallot, "Integration of depth modules: Stereo and shading," *J. Opt. Soc. Amer. A*, vol. 5, no. 10, pp. 1749–1758, 1988.

[8] M. S. Banks and B. T. Backus, "Extra-retinal and perspective cues cause the small range of the induced effect," *Vis. Res.*, vol. 38, no. 2, pp. 187–194, 1998.

[9] G. H. Recanzone, "Auditory influences on visual temporal rate perception," *J. Neuriophysiol.*, vol. 89, pp. 1078–1093, 2003.

[10] J. P. Bresciani, M. O. Ernst, K. Drewing, G. Bouyer, V. Maury, and A. Kheddar, "Feeling what you hear: Auditory signals can modulate tactile tap perception," *Exp. Brain Res.*, vol. 162, pp. 172–180, 2005.

[11] N. W. Roach, J. Heron, and P. V. McGraw, "Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration," *Proc. Roy. Soc. B: Biol. Sci.*, vol. 273, pp. 2159–2168, 2006.

[12] K. P. Kording, U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum, and L. Shams, "Causal inference in multisensory perception," *PLoS One*, vol. 2, no. 9, 2007, Art. no. e943.

[13] M. A. Srinivasan, G. L. Beauregard, and D. L. Brock, "The impact of visual information on the haptic perception of stiffness in virtual environments," in *Proc. ASME Int. Mech. Eng. Congr. Expo.*, 1996, pp. 555–559.

[14] M. Samad, E. Gatti, A. Hermes, H. Benko, and C. Parise, "Pseudo-haptic weight: Changing the perceived weight of virtual objects by manipulating control-display ratio," in *Proc. Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–13.

[15] M. Kuschel, M. D. Luca, M. Buss, and R. L. Klatzky, "Combination and integration in the perception of visual-haptic compliance information," *IEEE Trans. Haptics*, vol. 3, no. 4, pp. 234–244, Oct.–Dec. 2010.

[16] C. Basdogan, B. Ataseven, and M. A. Srinivasan, "Perception of soft objects in virtual environments under conflicting visual and haptic cues," *IEEE Trans. Haptics*, vol. 17, no. 2, pp. 227–236, Apr.–Jun. 2024.

[17] Y. Ujitoko and Y. Ban, "Survey of pseudo-haptics: Haptic feedback design and application proposals," *IEEE Trans. Haptics*, vol. 14, no. 4, pp. 699–711, Oct.–Dec. 2021.

[18] A. Paljic, J. M. Burkhardt, and S. Coquillart, "Evaluation of pseudo-haptic feedback for simulating torque: A comparison between isometric and elastic input devices," in *Proc. Int. Symp. Haptic Interfaces Virtual Environ. Teleoperator Syst.*, 2004, pp. 216–223.

[19] M. Tatezono, K. Sato, K. Minamizawa, H. Nii, N. Kawakami, and S. Tachi, "Effect of haptic feedback on pseudo-haptic feedback for arm display," in *Proc. IEEE ICROS-SICE Int. Joint Conf.*, 2009, pp. 4332–4337.

[20] M. Hosseini, A. Sengul, Y. Pane, J. D. Schutter, and H. Bruyninckx, "Haptic perception of virtual spring stiffness using ExoTen-glove," in *Proc. Int. Conf. Hum. Syst. Interact.*, 2018, pp. 526–531.

[21] E. Fakhoury, P. Culmer, and B. Henson, "The impact of visual cues on haptic compliance discrimination using a pseudo-haptic robotic system," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2017, pp. 1719–1725.

[22] D. Matsumoto, K. Hasegawa, Y. Makino, and H. Shinoda, "Displaying variable stiffness by passive nonlinear spring using visuo-haptic interaction," in *Proc. 2017 IEEE World Haptics Conf.*, 2017, pp. 587–592.

[23] K. v. Mensvoort, D. J. Hermes, and M. v. Montfort, "Usability of optically simulated haptic feedback," *Int. J. Hum.-Comput. Stud.*, vol. 66, no. 6, pp. 438–451, 2008.

[24] F. Argelaguet, D. A. G. Jáuregui, M. Marchal, and A. Lécuyer, "A novel approach for pseudo-haptic textures based on curvature information," in *Proc. EuroHaptics Conf.*, 2012, pp. 1–12.

[25] S. A. Klein, "Measuring, estimating, and understanding the psychometric function: A commentary," *Percep. Psychophysics*, vol. 63, no. 8, pp. 1421–1455, 2001.

[26] M. R. Leek, "Adaptive procedures in psychophysics research," *Percep. Psychophysics*, vol. 63, no. 8, pp. 1279–1292, 2001.

[27] E. Brochu, V. M. Cora, and N. d. Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," Dept. Comput. Sci., Univ. British Columbia, Vancouver, BC, Canada, Tech. Rep. UBC TR-2009-23, 2009. [Online]. Available: https://www.cs.ubc.ca/~nando/papers/bayopt.pdf

[28] A. B. Watson and D. G. Pelli, "Quest: A Bayesian adaptive psychometric method," *Percep. Psychophysics*, vol. 33, no. 2, pp. 113–120, 1983.

[29] L. L. Kontsevich and C. W. Tyler, "Bayesian adaptive estimation of psychometric slope and threshold," *Vis. Res.*, vol. 39, no. 16, pp. 2729–2737, 1999.

[30] N. Prins, "The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less)," *J. Vis.*, vol. 13, no. 7, pp. 1–17, 2013, doi: 10.1167/13.7.3.

[31] A. B. Watson, "QUEST+ : A general multidimensional Bayesian adaptive psychometric method," *J. Vis.*, vol. 17, no. 3, pp. 1–27, 2017, doi: 10.1167/17.3.10.

[32] N. Houlsby, F. Huszár, Z. Ghahramani, and M. Lengyel, "Bayesian active learning for classification and preference learning," 2011, *arXiv:1112.5745*.

[33] B. Settles, "Active learning literature survey," University of Wisconsin, Madison, WI, USA, Tech. Rep. TR1648, 2009.

[34] J. R. Gardner, X. Song, K. Q. Weinberger, D. L. Barbour, and J. P. Cunningham, "Psychophysical detection testing with Bayesian active learning," in *Proc. 31st Conf. Uncertainty Artif. Intell.*, 2015, pp. 286–295.

[35] L. Owen, J. Browder, B. Letham, G. Stocek, C. Tymms, and M. Shvartsman, "Adaptive nonparametric psychophysics," 2021, *arXiv:2104.09549*. [Online]. Available: https://arxiv.org/abs/2104.09549

[36] B. Letham, P. Guan, C. Tymms, E. Bakshy, and M. Shvartsman, "Look-ahead acquisition functions for Bernoulli level set estimation," in *Proc. Int. Conf. Artif. Intell. Statist.,* 2022, pp. 8493–8513.

[37] W. Felt, J. C. Selinger, J. M. Donelan, and C. D. Remy, ""Body-in-the-loop": Optimizing device parameters using measures of instantaneous energetic cost," *PLoS One*, vol. 10, no. 8, pp. 1–21, 2015.

[38] J. R. Koller, D. H. Gates, D. P. Ferris, and C. D. Remy, ""Body-in-the-loop" optimization of assistive robotic devices: A validation study," in *Proc. Robot.: Sci. Syst.*, 2015, pp. 1–10.

[39] Y. Ding et al., "Biomechanical and physiological evaluation of multi-joint assistance with soft exosuits," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 2, pp. 119–130, Feb. 2017.

[40] J. Zhang et al., "Human-in-the-loop optimization of exoskeleton assistance during walking," *Sci. Robot.*, vol. 356, no. 6344, pp. 1280–1284, 2017.

[41] X. Song, K. Sukesan, and D. Barbour, "Bayesian active probabilistic classification for psychometric field estimation," *Attention, Percep., Psychophysics*, vol. 80, pp. 798–812, 2018.

[42] E. Novoseller, "Online learning from human feedback with applications to exoskeleton gait optimization," Ph.D. dissertation, CalTech, Pasadena, CA, USA, 2021.

[43] C. Williams and D. Barber, "Bayesian classification with Gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1342–1351, Dec. 1998.

[44] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, vol. 7, no. 5, Cambridge, MA, USA: MIT Press, 2005.

[45] W. Chu and Z. Ghahramani, "Preference learning with Gaussian processes," in *Proc. Int. Conf. Mach. Learn.*, 2005, pp. 137–144.

[46] J. Schlittenlacher, R. Turner, and B. Moore, "Audiogram estimation using bayesian active learning," *J. Acoust. Soc. Amer.*, vol. 144, pp. 421–430, 2018.

[47] J. Browder, S. Bochereau, F. v. Beek, and R. King, "Stiffness in virtual contact events: A non-parametric Bayesian approach," in *Proc. IEEE World Haptics Conf.*, 2019, pp. 515–520.

[48] M. Tucker et al., "Human preference-based learning for high-dimensional optimization of exoskeleton walking gaits," in *Proc. IEEE Int. Conf. Robot. Syst.*, 2020, pp. 3423–3430.

[49] E. Biyik, N. Huynh, M. Kochenderfer, and D. Sadigh, "Active preference-based Gaussian process regression for reward learning," in *Proc. Robot.: Sci. Syst.*, 2020. [Online]. Available: https://www.roboticsproceedings.org/rss16/p041.pdf

[50] K. Li et al., "ROIAL: Region of interest active learning for characterizing exoskeleton gait preference landscapes," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 3212–3218.

[51] M. Tucker et al., "Preference-based learning for exoskeleton gait optimization," in *Proc. 2020 IEEE Int. Conf. Robot. Automat.*, 2020, pp. 2351–2357.

[52] B. Catkin and V. Patoglu, "Preference-based human-in-the-loop optimization for perceived realism of haptic rendering," *IEEE Trans. Haptics*, vol. 16, no. 4, pp. 470–476, Oct.–Dec. 2023.

[53] S. Lu, M. Zheng, M. C. Fontaine, S. Nikolaidis, and H. Culbertson, "Preference-driven texture modeling through interactive generation and search," *IEEE Trans. Haptics*, vol. 15, no. 3, pp. 508–520, Jul.–Sep. 2022.

[54] H. Z. Tan, N. I. Durlach, G. L. Beauregard, and M. A. Srinivasan, "Manual discrimination of compliance using active pinch grasp: The roles of force and work cues," *Percep. Psychophysics*, vol. 57, no. 4, pp. 495–510, 1995.

[55] L. Jones and I. Hunter, "A perceptual analysis of stiffness," *Exp. Brain Res.*, vol. 79, pp. 150–156, 1990.

[56] J. M. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy, "Combining sensory information: Mandatory fusion within, but not between, senses," *Science*, vol. 298, no. 5598, pp. 1627–1630, 2002.

[57] W. Richards, "Quantifying sensory channels: Generalizing colorimetry to orientation and texture, touch, and tones," *Sensory Processes*, vol. 3, pp. 207–229, 1979.

Harun Tolasa (Graduate Student Member, IEEE) received the B.Sc. degree in mechanical engineering from Bilkent University Ankara, Türkiye, in 2021, and the M.Sc. degree in mechatronics engineering from Sabanci University Istanbul, Türkiye, in 2024, where he is currently working toward the Ph.D. degree. His research interests include active learning, human-in-the-loop optimization, and haptic rendering.

Bilal Catkin received the B.Sc. degree in controls and automation engineering from Istanbul Technical University, Istanbul, Türkiye, in 2020, and the M.Sc. degree in mechatronics engineering from Sabanci University, Istanbul, in 2023. His research interests include optimization and physical human-robot interaction.

Volkan Patoglu (Member, IEEE) received the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor MI, USA, in 2005. He was a Postdoctoral Researcher with Rice University, Houston, TX, USA, in 2006. He is currently a Full Professor in mechatronics engineering with Sabanci University, Istanbul, Türkiye. His research focuses on the area of physical human-machine interaction, in particular, design and control of force feedback robotic systems with applications to rehabilitation. His research extends to cognitive robotics. He was an Associate Editor for IEEE TRANSACTIONS ON HAPTICS (2013–2017), IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING (2018–2023), and IEEE ROBOTICS AND AUTOMATION LETTERS (2019–2024).