

<https://doi.org/10.1038/s44182-026-00077-0>

Embodied tactile perception of soft objects properties

Anirvan Dutta¹, Alexis WM Devillard¹, Zhihua Zhang¹, Xiaoxiao Cheng² & Etienne Burdet¹ ✉

To enable robots to perform human-like dexterous manipulation, it is essential to understand how mechanical compliance, multi-modal sensing, and purposeful interaction jointly shape tactile perception. In this study, we use a dedicated *modular e-Skin* with interchangeable mechanical compliance and multi-modal sensing to systematically investigate how sensing embodiment and interaction strategies influence robotic perception of objects. Leveraging a curated set of soft *wave objects* with controlled viscoelastic and surface properties, we explore a rich set of *palpation primitives* that vary in indentation depth, frequency, and directionality. In addition, we propose the *latent filter*, an unsupervised, action-conditioned deep state-space model of the sophisticated interaction dynamics, and infer causal mechanical properties into a structured latent space. This provides in-depth, interpretable representation of how embodiment and interaction determine and influence perception. Our investigation demonstrates that multi-modal sensing outperforms unimodal sensing, emphasizing complex interaction between the environment and the mechanical properties of e-Skin.

Humans exhibit a remarkable ability for perceiving and manipulating various objects by performing exploratory actions like squeezing, palpating, and probing. The underlying sensory abilities are largely attributed to the multilayered, adaptive structure of human skin, populated with dense arrays of specialized mechanoreceptors¹. This sensorized skin, together with purposefully deployed exploration strategies, enables the extraction of rich, high-fidelity information during dynamic interactions, where both the skin and the contacted object deform^{2,3}. However, the specific contribution of the skin's mechanical properties, and the multi-modal sensing or *embodiment*, to tactile perception remains poorly understood⁴. Here, *embodiment* refers to how the physical body and its sensory apparatus shape perception and cognition, rather than the notion of agency and self-perception⁵.

Understanding how the mechanical properties of the skin shape the sensory signals and perception is crucial for robotic systems equipped with tactile sensing to achieve the perceptual capabilities of their biological counterparts. Studies by Costi et al.^{6,7} show that adding a soft outer layer to a force-sensing layer improves contact fidelity, but dampens high-frequency signals important for sensing surface texture—effectively acting as a mechanical low-pass filter. Follow-up work has explored how tuning this filter improves performance for specific object categories⁸, with softer skins better discriminating compliant materials and stiffer ones suited for rigid objects. Yet, these studies mostly focus on uniform outer layers and do not investigate the influence of multi-modal sensing and a multilayered mechanical structure. In this article, we systematically investigate embodied tactile perception by considering the sensory and mechanical aspects of an

artificial skin, systematic variations in the environmental surface and mechanical properties, and a repertoire of palpation primitives, as illustrated in Fig. 1. To this end, we developed a modular *electronic skin* (e-Skin)⁹ (see “Rich object *wave objects set*” section), which allows for controlled modulation of mechanical properties and sensing modalities—enabling systematic exploration of their influence on perception.

Previous studies have largely relied on simplified environments, such as biological tissue phantoms^{10–12} or homogeneous viscoelastic materials with basic geometries^{6,13}. While valuable, these configurations do not capture the diversity of naturally occurring soft objects, which often exhibit complex combinations of bulk mechanical properties and surface textures that jointly shape the perception of softness^{14,15}. To emulate this variability, we curated a foundational dataset of soft objects – *wave objects* (see “Rich object *wave objects set*” section) featuring controlled variations in bulk compliance and surface characteristics, enabling a comprehensive evaluation of embodied tactile perception.

In addition, the nature of exploratory interaction is critical to tactile perception. Human studies have shown that individuals adapt their exploration strategies—from simple probing to complex palpation—depending on the perceptual objective¹⁶. In contrast, robotic studies often rely on fixed, static probing motions¹⁷, overlooking how interaction parameters such as indentation depth, speed, and frequency affect perception. Moreover, sensory processing typically neglects temporal interaction dynamics, relying instead on static analyses that fail to capture how perception evolves during interaction. Without a systematic investigation of

¹Department of Bioengineering, Imperial College of Science, Technology and Medicine, London, UK. ²Department of Electrical & Electronic Engineering, The University of Manchester, Manchester, UK. ✉e-mail: e.burdet@imperial.ac.uk

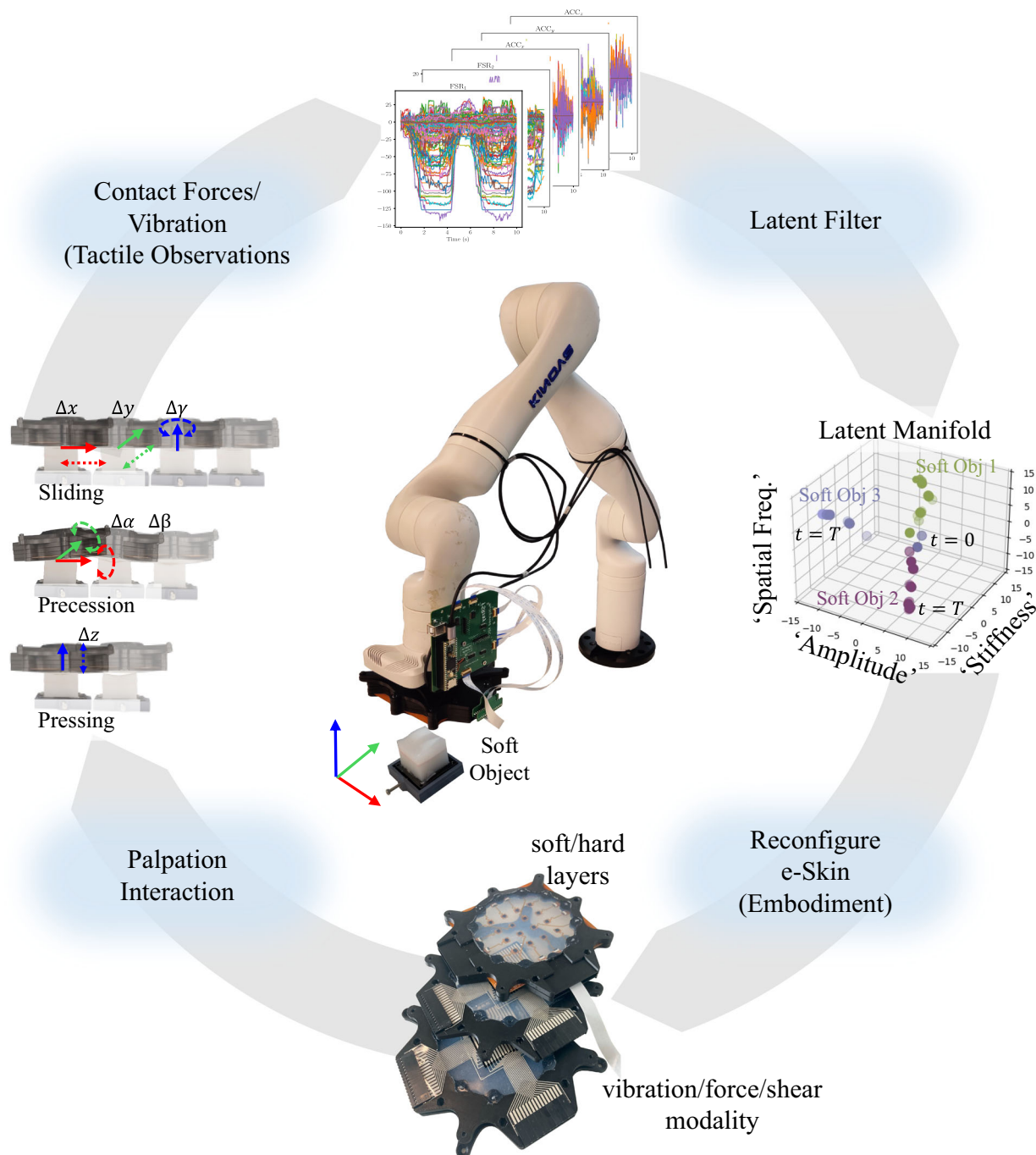


Fig. 1 | An interactive robotic setup to investigate how skin sensing and embodiment influence tactile sensing.

how dynamic, action-dependent exploration shapes tactile signals¹⁸, it remains difficult to develop a generalizable understanding of how embodiment and interaction jointly influence perception across diverse soft objects. To address this, we employ a repertoire of palpation-based interactions—including pressing, precession, and sliding, inspired by human soft-object exploration¹⁹ to elicit rich tactile responses during contact (see “Palpation interaction” section), enabling precise identification of salient mechanical properties while accounting for interaction dynamics²⁰.

Beyond these limitations in sensing and interaction, a central challenge lies in modeling soft sensor-object interaction. Most prior approaches are discriminative, relying on handcrafted statistical features^{21,22}, dimensionality reduction, or sophisticated classifiers to detect stiffness levels²³. Some works have applied techniques such as functional principal component analysis²⁴ or used kinesthetic sensing alone^{25,26}. While deep learning approaches—

using visuo-haptic CNNs^{13,27,28} or recurrent networks for temporal tactile signals^{29,30}—have shown improved performance, they are typically restricted to predefined object categories and lack physically grounded causal inference. Moreover, their performance is often highly sensitive to the specific soft-object set used during training and evaluation, highlighting the need for principled, regression-based models that support robust and unbiased analysis. Although regression-based approaches such as Contact Area Spread Rate (CASR) for bulk viscosity estimation³¹ or spring-damper models with Kalman filtering³² offer an alternative, they often depend on strong simplifying assumptions (e.g., material homogeneity or constrained motions) and struggle to generalize across diverse object types. These methods are further limited by their reliance on low-dimensional kinesthetic signals and by their exclusion of high-dimensional tactile information. Standardized metrics for soft-object characterization (e.g., stiffness, loss

tangent, Shore hardness) are similarly constrained to quasi-static, unimodal interaction paradigms³³. To overcome these challenges, we introduce an unsupervised, action-conditioned variational inference approach based on deep state-space modeling – *Latent Filter* (see “Deep state-space model: latent filter” section). This model captures the temporal dynamics of sensor-object interactions and learns a compact latent representation that encodes the underlying mechanical properties of soft objects, enabling generalizable and physically meaningful inference.

Through this study, we make the following key contributions toward a systematic understanding of embodied tactile perception by jointly considering the roles of mechanics, sensing, and interaction.

- We build on our prior work to leverage a first-of-its-kind modular *electronic skin* capable of seamless controlled modulation of mechanical properties and sensing modalities. This novel e-Skin enabled a systematic investigation of how embodiment shapes tactile perception, not explored previously.
- We propose *Latent Filter*, an unsupervised, action-conditioned variational inference framework that models complex skin-object interaction dynamics through a deep state-space formulation, inferring coherent, physically meaningful latent representations for detailed analyses of mechanics, sensing on perception.
- We design comprehensive experimental settings using a curated set of soft *wave objects* with varied bulk and surface properties, combined with diverse palpation-based robotic interactions, to evaluate embodied tactile perception across realistic conditions.

We performed extensive ablation studies to derive key insights reported in this work, while also comparing with baseline techniques.

Results

This section presents the outcomes of our study using the proposed framework. We first describe the modular *e-Skin*, which enables systematic variation of mechanical compliance and sensing modalities, and the curated set of soft objects with controlled viscoelastic and geometric properties. We then introduce the proposed *Latent Filter*—a deep, unsupervised, action-conditioned state-space model designed to infer intrinsic soft-object properties from temporal tactile observations. Leveraging this framework, we conduct a series of experiments to analyze how sensor embodiment, sensing modalities, and interaction strategies jointly influence the perception of soft-object properties.

e-Skin tactile sensor

The modular e-Skin developed for this study features a layered structure, with each layer composed of materials and sensors tailored to control mechanical properties and sensing capabilities. This modularity enables the replacement or reconfiguration of individual layers to suit different tasks and environments. In the configuration used here (Fig. 2a), the e-Skin comprises three sensorized layers embedded within compliant silicone. Two variants were fabricated by varying the silicone material: a soft e-Skin using Smooth-On Ecoflex (Shore 00-31) and a more rigid e-Skin using Dragon-skin (Shore 30A). Notably, the Shore A and Shore 00 scales measure different hardness ranges; DragonSkin’s value of 30A indicates a soft yet structurally stable elastomer, while Ecoflex 00-31, measured on the softer Shore 00 scale, corresponds approximately to Shore A 10-15.

The outermost layer houses a 4×4 array of multi-nodal 3-axis accelerometers, positioned just beneath a thin silicone layer to minimize damping and detect high-frequency vibrations—analogueous to rapidly adapting mechanoreceptors in human skin. To measure normal forces and deformations, two layers of *force-sensing resistors* (FSRs) were embedded, each forming a 16×16 array. FSRs are inherently limited to sensing normal forces, so a compliant inter-layer of silicone was introduced between the two arrays. Shear forces may be indirectly estimated by capturing differential normal force patterns across the FSR layers during shear displacements. This multilayered, multi-nodal configuration enables detailed examination of sensing modalities and mechanical compliance of artificial tactile sensing. As the signals across layers are intrinsically coupled, capturing the

underlying sophisticated skin-object interaction dynamics necessitates a robust modeling framework that can uncover the key causal properties governing the interaction³ and identify the characteristic features of the tactile sensorimotor interactions^{4,34}.

Rich object wave objects set

To evaluate the e-Skin’s performance in a controlled yet diverse setting of object properties, we designed a set of soft *wave objects* with parametrically controlled surface and stiffness properties (Fig. 2b). Each object was defined by three parameters: amplitude, spatial frequency (surface texture), and stiffness (bulk compliance). Surface textures were created by applying a band-pass filter to white noise images, centered at spatial frequencies {10/m, 30/m, 50/m}. This method generated surface profiles with natural variability while retaining control over key parameters. Amplitudes were scaled to one of three levels {5, 10, 20} mm. The mirrored surfaces were 3D printed to produce molds for silicone casting. Stiffness variation was achieved using different materials—Ecoflex 00-10, Ecoflex 00-50, and rigid PLA—each with distinct Shore hardness. To introduce heterogeneity, a subset of the objects included a stiff top layer (heterogeneity $\in \{0, 1\}$). This dataset captures a wide range of perception conditions relevant to soft-object manipulation and provides a foundation for systematic tactile perception evaluation.

Deep state-space model: latent filter

We model the interaction between soft objects and the robot equipped with e-Skin as a discrete nonlinear dynamical system with tactile observations (vibrations, contact forces) $\mathbf{o}_{1:T} = (\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T)$, $\mathbf{o}_t \in \mathcal{O} \subset \mathbb{R}^{n_o}$ in discrete time steps $t = 1, \dots, T$, and actions $\mathbf{a}_{1:T} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_T)$, $\mathbf{a}_t \in \mathcal{A} \subset \mathbb{R}^{n_a}$. These interactions induce complex spatiotemporal tactile observations that encode information about the object’s underlying mechanical properties. An unsupervised deep state-space model is employed to capture interaction dynamics by leveraging time-series and action-conditioned tactile data through low-dimensional latent variables $\mathbf{s}_{1:T} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_T)$ with $\mathbf{s}_t \in \mathcal{S} \subset \mathbb{R}^{n_s}$ that represent the underlying state of the system. The objective is to model the joint probability $p(\mathbf{o}_{1:T}, \mathbf{s}_{1:T} | \mathbf{a}_{1:T})$, and to perform variational inference by maximizing the likelihood of observations

$$p(\mathbf{o}_{1:T} | \mathbf{a}_{1:T}) = \int p(\mathbf{o}_{1:T}, \mathbf{s}_{1:T} | \mathbf{a}_{1:T}) d\mathbf{s}_{1:T} \quad (1)$$

We hypothesize that accurately modeling and inferring the dynamics of the robotic system during its interaction with soft objects will enable the extraction of causal physical factors underlying the process. To support this, we introduce a structural assumption in the latent space grounded in the concept of observability from classical control theory. Specifically, we distinguish between directly observable components of the latent state—those that can be inferred from a single observation, dependent on indirectly observable components—those that require observations accumulated over multiple time steps for accurate estimation. This facilitates analytical computation of the posterior while avoiding the computational expense of linearizing the observation model. We propose restructuring/partitioning of the latent state-space:

$$p(\mathbf{s}_t) = p(\mathbf{z}_t, \mathbf{y}_t) = p(\mathbf{z}_t | \mathbf{y}_t) p(\mathbf{y}_t) \quad (2)$$

with directly observable variable $\mathbf{z}_t \in \mathbb{R}^{n_z}$ and indirectly observable part as $\mathbf{y}_t \in \mathbb{R}^{n_y}$, $n_s = n_z + n_y$ with updated *Evidence Lower Bound Objective* (ELBO): (see the Supplementary Materials, Section 1 for detailed derivation):

$$\mathcal{F}_{\text{ELBO}}(\theta, \psi) = \mathbb{E}_{q_\psi(\cdot)} \left[\sum_{t=1}^T \log p_\theta(\mathbf{o}_t | \mathbf{z}_t) \right] - \sum_{t=2}^T \text{KL}[q^{\text{fill}}(\mathbf{z}_t) \| p(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{y}_t, \mathbf{a}_t)] \\ - \sum_{t=2}^T \text{KL}[q_\psi(\mathbf{y}_t | \cdot) \| p(\mathbf{y}_t | \mathbf{a}_t, \mathbb{N})] \quad (3)$$

The architecture of the proposed latent filter is illustrated in Fig. 3, where we employ the inverse variational measurement model $q_\psi(\mathbf{z}_t | \mathbf{o}_t)$ and

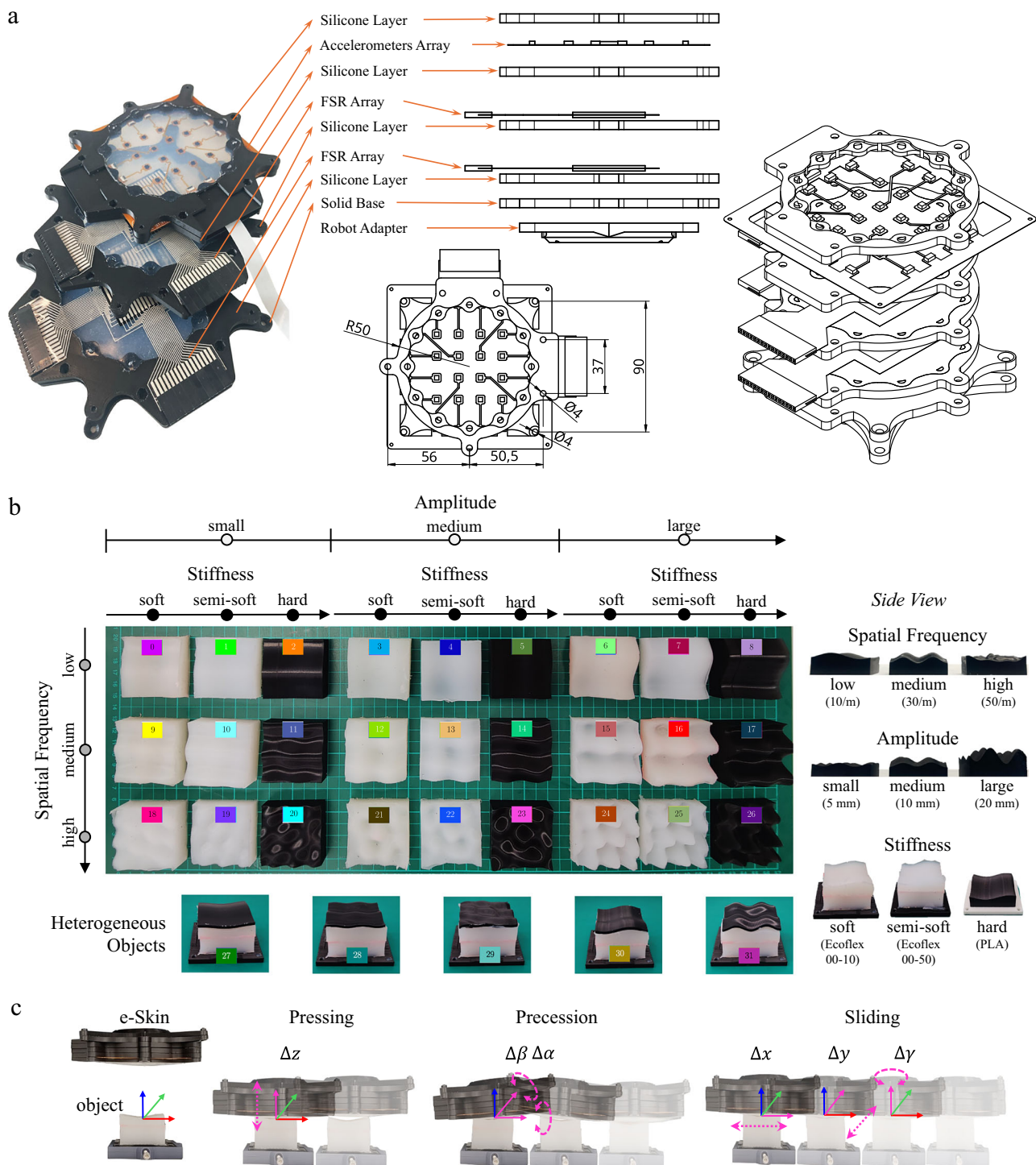


Fig. 2 | E-skin, object set, and movement primitives to investigate tactile perception for robotics. **a** Diagram of the modular e-Skin⁹ used in the current paper's study. **b** Wave objects with controlled surface and stiffness properties. Objects 0–26

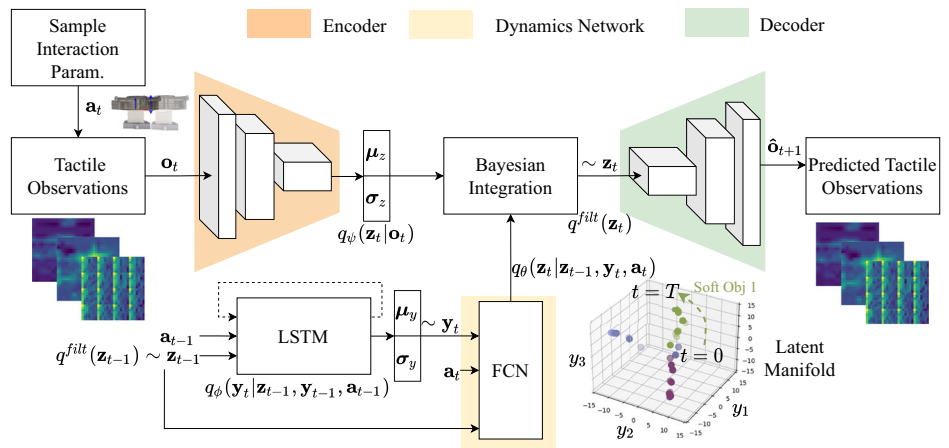
are homogeneous in material; objects 27–31 are heterogeneous, incorporating a thin, stiff top layer. **c** Robotic palpation primitives designed to elicit rich tactile responses. Active motion axes are shown in pink.

perform Bayesian integration with the recursive model $q_{\theta}(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{y}_t, \mathbf{a}_t)$ to compute the filtered variational distribution $q^{fil}(\mathbf{z}_t)$. To approximate the indirectly observable distribution $q_{\psi}(\mathbf{y}_t | \mathbf{z}_{t-1}, \mathbf{y}_{t-1}, \mathbf{a}_{t-1})$, we utilize an LSTM network. Additionally, we incorporate a learnable hierarchical prior $p(\mathbf{y}_t | \mathbf{a}_t, \mathbb{N})$, where $\mathbb{N}$ represents the soft-object label, to enforce the extraction of causal physical properties while relaxing constraints on time sequence length³⁵. The complete latent filter model is optimized end-to-end using the Evidence Lower Bound (ELBO) objective defined in Eq. (3).

Experiments

We conducted a systematic study examining how sensor embodiment, sensing modality, and interaction strategy jointly influence the perception of soft-object properties. Specifically, we investigated the effects of (i) e-Skin stiffness with soft—Ecoflex vs. hard—DragonSkin, (ii) individual sensing modalities, including the accelerometer layer and the two force-sensing resistor (FSR) layers (top: FSR_t, bottom: FSR_b), and (iii) three palpation primitives—pressing, precession, and sliding (see “Palpation interaction”

Fig. 3 | Proposed latent filter architecture. The encoders implement the inverse variational measurement model and are fused via Bayesian integration with the dynamics model, approximated by a fully connected network (FCN). The decoder, mirroring the encoder structure, reconstructs tactile observations. Encoder architectures vary depending on the input modality combination. An LSTM (Long Short-Term Memory) network estimates the indirectly observable latent distribution.



section), on the perception of key object properties, namely spatial frequency, amplitude, and stiffness as shown in Figure 2b, c. An initial experimental evaluation was performed to identify suitable parameter ranges considering the robot controller frequency and sensor saturation. Within these bounds, 16 uniformly sampled values were selected to perform interaction actions and record tactile observation data. Each trial began with the robot moving to a fixed position, descending until a predefined contact force was reached, and then executing the palpation trajectory for 10 s. A total of 2560 interaction trajectories were collected, each comprising 7000 samples (sampling frequency: 700 Hz) for both e-Skin stiffness configurations. An illustrative plot of the raw tactile observation is presented in Figure 1 (see Supplementary Material). The raw e-Skin observation was processed to form the tactile observations prior to model training, the details of which are presented in “Data processing and training” section. Tactile observations $\mathbf{o}_{1:T}$ included two-layer FSR and accelerometer readings with dimensionality $n_o = 28 \times 28 \times 3$. The action observations are denoted as $\mathbf{a}_{1:T} = [\mathbf{p}_t, \boldsymbol{\theta}_t, \mathbf{p}_t, \boldsymbol{\omega}_t]_{t=1}^T$, where $\mathbf{p}_t \in \mathbb{R}^3$ represents the end-effector position (x, y, z), $\boldsymbol{\phi}_t \in \mathbb{R}^3$ represents the orientation in Euler angles, $\mathbf{p}_t \in \mathbb{R}^3$ is the linear velocity, and $\boldsymbol{\Omega}_t \in \mathbb{R}^3$ is the angular (rotational) velocity, all in global co-ordinates, yielding a total dimensionality of $n_a = 12$.

To evaluate the effectiveness of our proposed *Latent Filter* in extracting physically meaningful features, we compared it against three state-of-the-art unsupervised learning techniques. Baseline I employed a static VAE model adapted from ref. 36. Baseline II is a sequential VAE (sVAE)³⁷, and Baseline III was an ablation of the proposed *Latent Filter* without the learnable hierarchical prior. In addition, we compared two multi-modal fusion strategies (early and late fusion) to determine how different sensory observations should be integrated to optimize perceptual performance (see “Data processing and training” section, Fig. 9b). The following sections present the results obtained across varying e-Skin configurations and palpation strategies.

Analysis of latent space

The learned latent features were systematically evaluated. Figure 4a presents 3D UMAP³⁸ embeddings of these latent representations (last time step, post-convergence), comparing the proposed approach with the baseline methods. The visualization clearly shows that the baseline techniques fail to learn meaningful features that could enable object-level separation. Interestingly, Baseline III exhibits partial clustering behavior, although the separation is notably weaker than that achieved by the proposed model. We hypothesize that careful tuning of the regularization parameters could further improve its performance; however, the hierarchical prior in our approach inherently mitigates the need for such fine balancing.

Whereas Fig. 5a presents 3D UMAP³⁸ embeddings for different modality configurations: ACC (accelerometer only), FSR_{tb} (both layers of FSR), MULTI-L (soft) (multi-modal late fusion with soft skin), and MULTI-

L (hard) (multi-modal late fusion with hard skin) across the three interaction primitives. These UMAP plots reveal that multi-modal configurations (MULTI-L), particularly with soft skin, exhibit the most structured and separable clusters, indicating their ability to encode discriminative features. FSR_{tb} alone also produces distinct clusters for many object classes, especially in precession and pressing interactions, underscoring the informativeness of 2-layer contact force information. In contrast, embeddings from the ACC-only setting appear more entangled and dispersed, suggesting that transient observations are insufficient for capturing object-specific features. In addition, it also demonstrates that the combination of bulk exploration via pressing and precession provides more discriminative tactile observations than surface-level interaction of sliding. As UMAP projections do not preserve relative spatial relationships from the original high-dimensional latent space, both intra-class and inter-class Euclidean distances were computed using object labels to provide a complementary quantitative assessment of latent space separability and confusion among the inferred soft-object features. Figure 5b and 4b present the distances normalized to [0,1] as heatmaps.

These heatmaps confirm the UMAP observation: Baseline approaches, with the exception of Baseline III, do not capture any meaningful features, and MULTI-L (soft), particularly under precession, achieves superior inter-class separability and reduced confusion. The ACC-only modality consistently shows poor discrimination and high confusion specially during precession interaction where consistent stable contact was made. However, ACC can detect features by sliding that FSRs alone cannot distinguish. An intriguing observation is that the MULTI-L (Hard) configuration exhibits increased confusion for object classes 12–26, which are characterized by complex surface features (higher surface spatial frequency). This suggests that the rigid skin configuration may limit the tactile sensor’s ability to capture fine surface details critical for class separation. However, Euclidean distance metric in such spaces may not be a faithful proxy for true dissimilarity, especially if the latent manifold is non-Euclidean and influenced by hierarchical variational priors. Furthermore, normalization of distances could obscure subtle yet meaningful inter-class distinctions. We further conducted a quantitative evaluation of the learned latent representations in terms of both clustering and classification performance, as summarized in Figure 3 (see Supplementary Materials). Specifically, classification accuracy was assessed using a logistic regression model, while clustering quality was quantified using the Silhouette L2 score³⁹ and the Davies–Bouldin score⁴⁰.

As the synthetic object set is parametrized by known mechanical properties (e.g., stiffness, spatial frequency, amplitude), a more principled analysis was conducted by aligning the latent space with these ground-truth generative factors, using the supervised regression method proposed in ref. 41. This allowed investigation of the time-series evolution of physically meaningful latent features, providing in-depth analysis for this study that extends beyond discrete class boundaries.

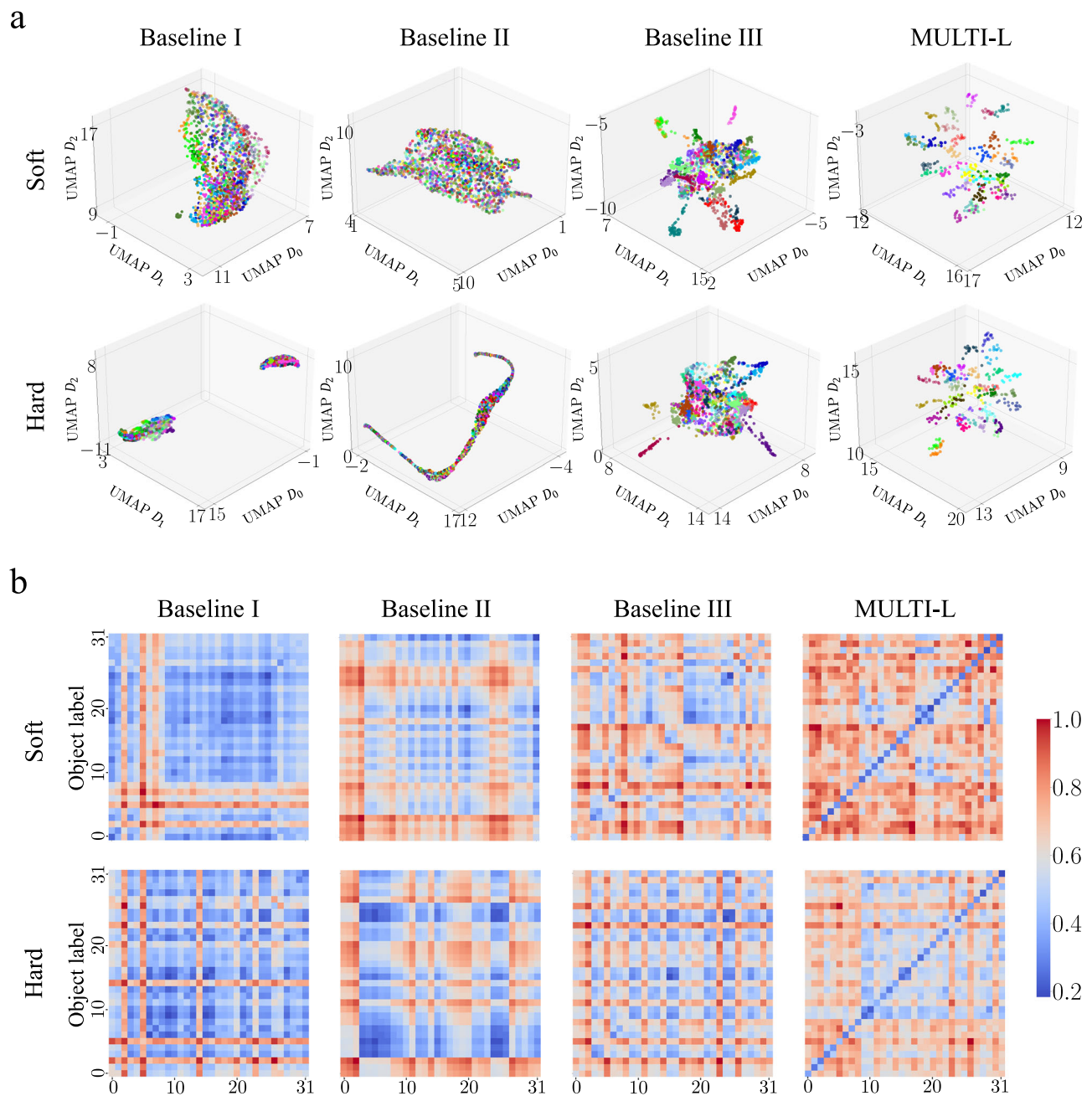


Fig. 4 | Comparison of latent representations across different baseline techniques and the proposed approach. a UMAP projections highlight cluster separability, with each unique color signifying a soft object. **b** Distance matrices quantify intra-class and inter-class similarity in the latent space with colors scaled to min-max values.

Analysis on multi-modality of e-Skin

To quantitatively relate the learned latent features to the underlying physical properties, a nonlinear kernel ridge regression model⁴² was employed. The final segments of each interaction sequence were used to train the regressor, which was then applied to predict the underlying mechanical properties of the synthetic objects—stiffness, surface spatial frequency, amplitude, and material heterogeneity of samples drawn from the evolving latent space. Example plots of two instances of filtering following alignment are depicted in the Supplementary Material (Figure 2). To assess alignment performance, the normalized mean squared error (NMSE) was computed between the predicted mechanical properties and ground truth. Normalization was performed to account for the differing numerical ranges and units associated with each mechanical property, ensuring fair comparison across metrics. Figure 6 presents the temporal evolution of the NMSE, comparing different combinations of modality across the three primitives. The

evaluated modalities include: ACC (accelerometer only), FSR₁ (first FSR layer), FSR₂ (second FSR layer), FSR₁₂ (both FSR layers), MULTI-E (early fusion of accelerometer and both FSR layers), and MULTI-L (late fusion of accelerometer and both FSR) (see Figure 9b for early and late fusion). For consistency, all of these evaluations were performed using the soft skin. Analyzing the time-series of the prediction error enables assessment of how efficiently each modality combination supports inference. This information is lost if NMSE were averaged over the entire sequence or with other static analyses done in prior works.

The results demonstrate that the late-fusion multi-modal approach (MULTI-L) consistently yields the lowest NMSE across all interaction primitives, indicating more robust and discriminative representations. In addition, we can also observe the effect of two-layer FSR in estimating stiffness and heterogeneity property, highlighting the importance of capturing shear force or skin-stretch information. In addition, transient

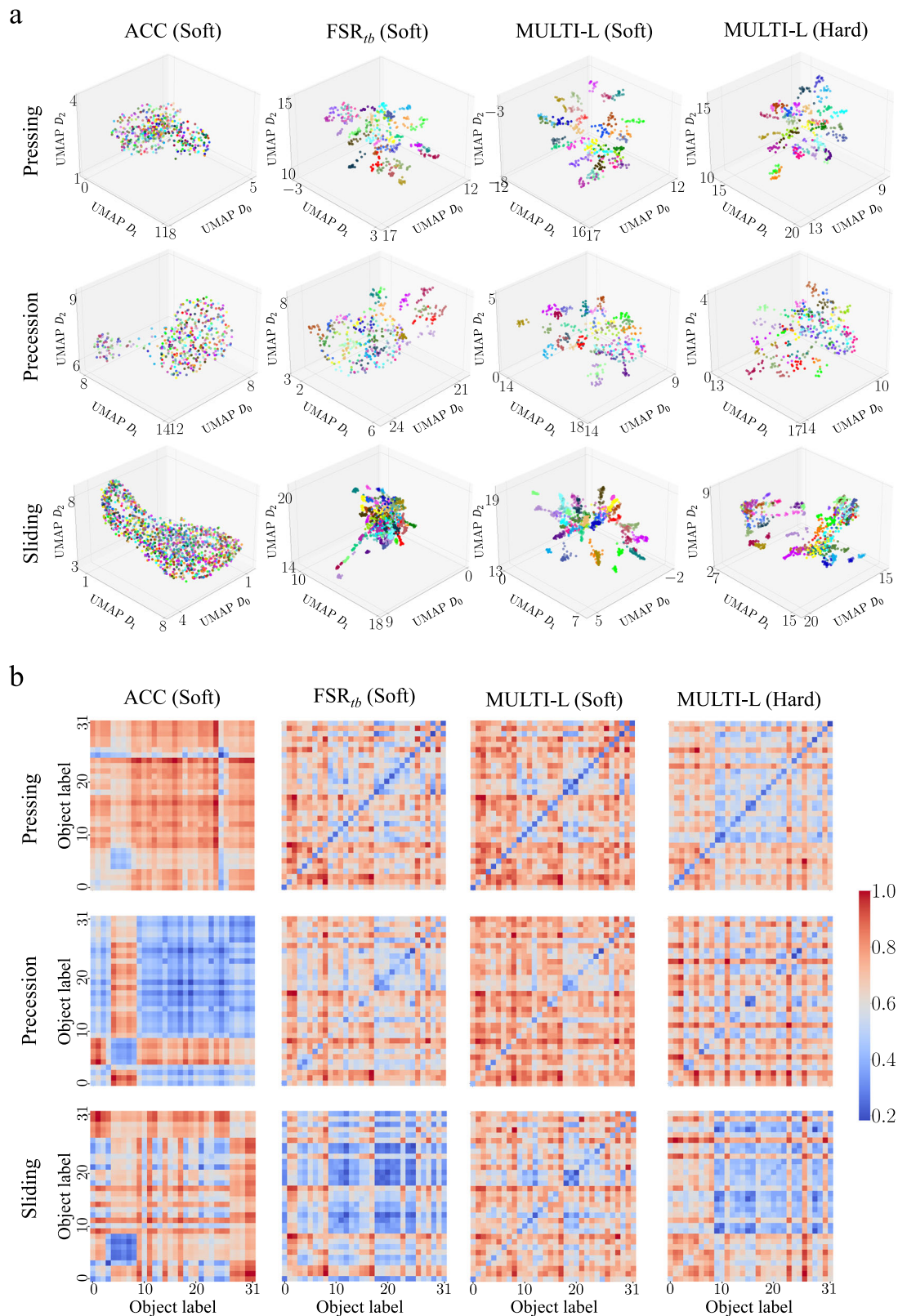


Fig. 5 | Comparison of latent representations across different tactile sensor modalities, skin stiffness conditions, and interaction primitives. a UMAP projections highlight cluster separability, with each unique color signifying a soft object.

b Distance matrices quantify intra-class and inter-class similarity in the latent space with colors scaled to min-max values.

accelerometer signals contribute to faster, albeit less accurate inference, and when combined with a two-layer FSR, improve the performance when performing surface-level exploration of sliding. Interestingly, although the early fusion scheme (MULTI-E) aligns more closely with the physical nature

of sensor co-activation, it results in less consistent performance. This discrepancy reveals a fundamental mismatch in the nature of force and vibration signals. While force data tends to be smooth and contact-driven, the spectrogram of the vibration signals results in a distinct observation

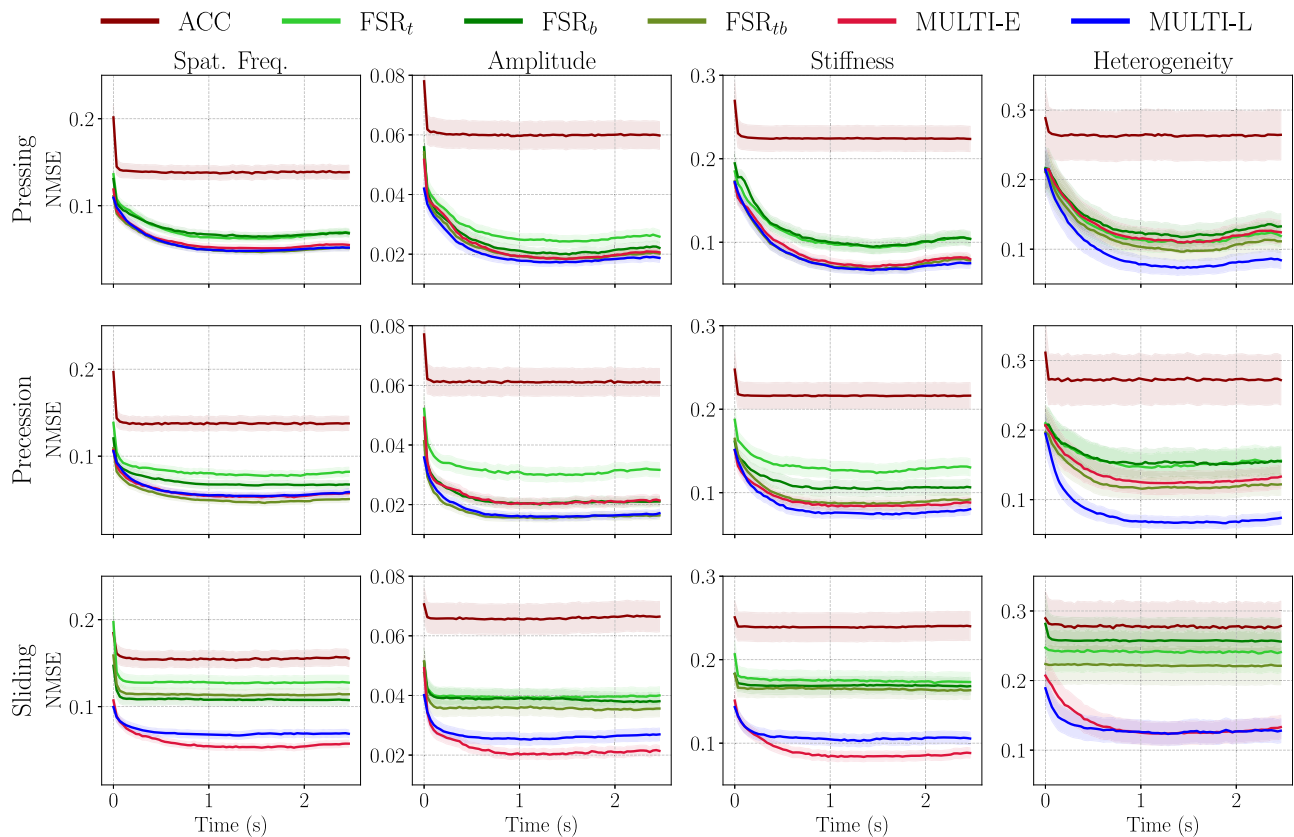


Fig. 6 | Detailed analysis of the different combinations and study of the multi-modality for soft skin. Standard deviation is scaled $\times 10$ for visual clarity.

distribution, indicating that separate encoding pathways are better suited for capturing their respective distribution. These findings highlight not only the need to evaluate multi-modality but also the importance of developing modality-aware tactile encoders capable of disentangling heterogeneous signal sources. Future research could focus on the design of such tactile-specific encoders, enabling more efficient and interpretable perception in multi-modal artificial e-Skins.

Analysis of interaction

The regression approach on the aligned latent values was further applied to investigate the influence of interaction parameters on the perception of soft objects across different components and interaction types. Figure 7 lists all the 16 parameters used for the interactions along with their corresponding NMSE values computed over the entire interaction trajectory. We also report the temporal evolution of the NMSE for best and worst performing parameters, selected to avoid visual and interpretive clutter in Figure 4 (see Supplementary Material).

The results reveal that interactions involving moderately higher frequencies (~ 0.6 Hz) and greater contact depths lead to better parameter inference across all interaction types. However, performance deteriorates at the highest tested frequencies, suggesting the existence of an optimal frequency range for effective palpation of soft objects. The influence of interaction parameters is particularly pronounced in precession and sliding, where they significantly affect estimation quality. In contrast, for pressing interactions, parameter variations primarily influence convergence speed rather than final estimation accuracy. Notably, in surface-level sliding, movement amplitude has a limited impact, with interaction frequency playing a more dominant role. This analysis provides key insights into the different nature of the palpation interaction for soft-object perception.

Analysis of the stiffness of e-Skin

An analysis of the influence of the e-Skin's mechanical stiffness on perception was conducted using the MULTI-L configuration (see Figure 5 in

Supplementary Material). The findings indicate that hard skin (Dragon-Skin-30) consistently yields better performance in estimating spatial frequency, amplitude, and bulk stiffness, whereas softer skin (Ecoflex 00-31) achieves improved accuracy when interacting with complex heterogeneous objects, particularly during sliding interactions. However, because the distribution of soft and hard objects in the dataset was not uniform, and to further explore whether perceptual performance is influenced by the mechanical similarity between the skin and the object⁶, we grouped the objects into three categories: soft, consisting of items with low to medium stiffness and small-to-medium surface amplitudes (Objects 0,1,3,4,9,10,12,13,18,19,21,22); heterogeneous (Objects 27–31); and hard, comprising objects with high stiffness or large surface amplitude/complexity. This grouping is motivated by human perceptual studies, which indicate that perceived softness is influenced by the viscoelastic properties of the object and its surface irregularities. NMSE errors were then recomputed for each group over the interaction sequence and are reported in Fig. 8.

This object-specific analysis reveals that soft e-Skin offers a perceptual advantage when interacting with soft objects, though this benefit is primarily observed during pressing interactions and is most prominent in the estimation of surface heterogeneity and bulk stiffness. This advantage is particularly evident during sliding and pressing interactions. However, this difference is diminished for precession interaction where stable and stronger contact was established between the e-Skin and the soft objects. Conversely, hard e-Skin performs more reliably across hard objects, maintaining consistent accuracy across interaction types. Notably, soft skin is particularly well-suited for perceiving the properties of heterogeneous materials with spatially varying viscoelasticity, as well as for surface-level interactions such as sliding. Overall, the results highlight that the perceptual effectiveness of a given e-Skin stiffness depends both on object properties and on the interaction type, suggesting the opportunities offered by co-designing the embodiment of tactile sensing together with exploration strategies.

Pressing			Precession			Sliding			
A_z (mm)	ω_z (Hz)	Traj. NMSE	$A_{\alpha,\beta}$ (mm)	$\omega_{\alpha,\beta}$ (Hz)	Traj. NMSE	$A_{x,y}$ (mm)	A_γ (°)	$\omega_{x,y,\gamma}$ (Hz)	Traj. NMSE
2.0	0.2	7.409	3.0	0.2	7.483	2.0	3.0	0.2	10.051
2.0	0.4	6.816	3.0	0.4	5.860	2.0	3.0	0.4	7.143
2.0	0.6	6.821	3.0	0.6	5.936	2.0	3.0	0.6	7.088
2.0	0.8	6.888	3.0	0.8	6.946	2.0	3.0	0.8	9.536
4.0	0.2	6.556	4.0	0.2	6.784	4.0	4.0	0.2	9.889
4.0	0.4	6.351	4.0	0.4	4.867	4.0	4.0	0.4	6.646
4.0	0.6	6.204	4.0	0.6	5.036	4.0	4.0	0.6	6.591
4.0	0.8	5.934	4.0	0.8	6.451	4.0	4.0	0.8	9.172
6.0	0.2	6.419	5.0	0.2	6.676	6.0	5.0	0.2	9.978
6.0	0.4	6.185	5.0	0.4	4.657	6.0	5.0	0.4	6.763
6.0	0.6	5.953	5.0	0.6	4.593	6.0	5.0	0.6	6.946
6.0	0.8	5.969	5.0	0.8	6.539	6.0	5.0	0.8	9.150
8.0	0.2	6.285	6.0	0.2	6.803	8.0	6.0	0.2	10.166
8.0	0.4	5.928	6.0	0.4	4.758	8.0	6.0	0.4	7.367
8.0	0.6	5.852	6.0	0.6	3.923	8.0	6.0	0.6	6.924
8.0	0.8	5.877	6.0	0.8	6.850	8.0	6.0	0.8	9.164

Fig. 7 | Result of action parameter with orange hue heatmap for each interaction (better performance indicated by lighter color). The traj. NMSE is the mean value computed of the entire interaction trajectory over time and is scaled by a factor of 10^{-2} .

		Soft Objects			Hard Objects			Heterogeneous Objects		
		Pressing	Precession	Sliding	Pressing	Precession	Sliding	Pressing	Precession	Sliding
Soft e-Skin	Spatial Freq.	0.022	0.025	0.031	0.066	0.072	0.082	0.055	0.055	0.070
	Amplitude	0.008	0.007	0.013	0.027	0.024	0.036	0.009	0.007	0.010
	Stiffness	0.032	0.047	0.052	0.112	0.108	0.134	0.023	0.030	0.070
	Heterogeneity	0.017	0.016	0.043	0.017	0.025	0.032	0.516	0.376	0.629
	Overall	0.017	0.023	0.035	0.052	0.057	0.070	0.130	0.105	0.191
Hard e-Skin	Spatial Freq.	0.021	0.014	0.040	0.052	0.047	0.096	0.045	0.048	0.080
	Amplitude	0.007	0.006	0.012	0.022	0.018	0.041	0.008	0.008	0.017
	Stiffness	0.042	0.044	0.070	0.088	0.084	0.125	0.031	0.047	0.094
	Heterogeneity	0.054	0.052	0.028	0.015	0.027	0.029	0.543	0.685	1.172
	Overall	0.032	0.029	0.037	0.040	0.042	0.072	0.130	0.176	0.339

Fig. 8 | Tabular result with Traj. NMSE values over the entire interaction sequence (lower is better), analyzing whether there is an inherent advantage of the properties of the object and the mechanical properties of e-Skin. The overall value is computed by taking the mean over all the dimensions of softness at the last time step, post-convergence.

Discussion

This paper introduces a principled framework for investigating how variations in interaction strategies and sensor embodiment influence tactile perception of soft and rigid objects. It presented three core contributions: (i) the development of a modular, multi-modal *e-Skin* tactile sensor with interchangeable mechanical compliance, capable of capturing both spatial force and vibration signals; (ii) the design of a foundational dataset of *wave objects* that systematically varies in viscoelasticity, surface texture, and heterogeneity, representing the diversity of natural environments; and (iii) the *Latent Filter*, an unsupervised, action-conditioned deep state-space model that encodes interaction dynamics into a structured latent manifold for causal inference of object properties.

Comparative analyses against baseline methods show that the proposed *Latent Filter* more effectively captures the complex relationships between sensing embodiment, interaction strategy, and sensing modality. Furthermore, the incorporation of a hierarchical prior further enhances performance relative to the ablative Baseline III. This prior is applied exclusively during training and does not restrict the generative model's ability to generalize during inference, in contrast to the rigid class conditioning employed in discriminative models.

The results of our study highlighted the importance of taking into account action parameters (action-conditioned modeling), which prior

approaches often overlook in their treatment of tactile interaction. We also observed that a single unified model did not generalize well across diverse interaction primitives. This limitation is likely due to the relatively small dataset compared to typical deep learning requirements, motivating us to train separate models for each interaction type. Although larger architectures, substantially more data, or hierarchical/context-aware priors may eventually support a shared model, investigating these directions lies beyond the scope of the present work.

Our findings from diverse interaction scenarios highlight several important aspects of tactile perception. High-dimensional, multi-modal sensing proved essential for generalization, as different modalities dominated in different interaction contexts. As in humans⁴³, the fusion of multiple modalities with the same sense offers benefits relative to each of the separate modalities. This inclusion of two FSR layers was particularly beneficial for estimating stiffness and heterogeneity, since differential normal forces provide a proxy for shear forces and local skin deformation. Although accelerometer-only embeddings were less discriminative overall, they captured valuable transient information during surface-level interactions, such as sliding—information that the FSR layers alone could not provide. The modular, layered design of the *e-Skin* opens new avenues for exploring fine-grained mechanical effects, such as the role of artificial

fingerprints, and enables robotic systems with enhanced tactile embodiment through both mechanical versatility and rich sensing capabilities.

Importantly, sensory processing with late fusion of different modalities consistently outperformed early fusion and unimodal approaches. This suggests that considering separate encoding pathways prior to integration can prevent representation mismatches that may arise when combining signals with different temporal and spatial statistics. These results indicate the need for modality-aware tactile encoders that can explicitly disentangle force – and vibration-based features, which is a promising direction for achieving robust and interpretable perception.

The motion interaction strategy also emerged as a critical factor shaping perception. By incorporating temporal dynamics, the latent filtering approach with the regression technique in “Analysis on multi-modality of e-Skin,” “Analysis on interaction,” and “Analysis on stiffness of e-Skin” sections achieved more interpretable and sensitive representations than static latent space analyses, such as UMAP or Euclidean distance metrics presented in “Analysis of latent space” section. Motion with moderate indentation frequencies (~0.6 Hz) and sufficient contact depth consistently resulted in low NMSE across properties, with sliding and precession interactions benefiting most from parameter tuning. In contrast, pressing was less sensitive to parameter variation, mainly affecting convergence speed rather than final estimation accuracy. These observations are consistent with findings from human haptic exploration, where perception is actively modulated by task-specific exploratory procedures⁴⁴. The results, therefore, suggest that robotic palpation strategies should be adaptively selected according to the property of interest and interaction strategy.

The mechanical stiffness of the e-Skin further influenced perceptual performance in a property- and interaction-specific manner. A stiffer skin (with DragonSkin) led to more accurate bulk property estimation due to improved force transmission, whereas a softer skin (with Ecoflex) enhanced the perception of heterogeneous or compliant objects, particularly in sliding interactions where surface-level features dominate. This suggests that the embodiment of tactile sensing could be co-optimized with interaction strategies rather than treated as a fixed design choice. In addition, the *Latent Filter* proved effective in extracting causal, interpretable representations of object properties by leveraging action-conditioned temporal dynamics. By explicitly modeling both directly and indirectly observable latent components, the approach bridges the gap between discriminative deep learning methods and analytical models, thereby supporting generalizable perception across diverse soft objects.

Our approach, which combines a modular *e-Skin* with the *Latent Filter* framework, provides a principled foundation for designing and customizing artificial tactile sensing for robotic applications such as medical tissue examination, food handling, and the manipulation of delicate or deformable objects. This work also underscores the importance of modeling tactile perception in a way that accounts for action/interaction and embodiment—factors long recognized in human studies but often overlooked in robotics. Overall, the study advances our understanding of robotic tactile perception and advocates for the development of dedicated methods for haptics rather than relying on approaches directly adapted from machine vision.

In future work, we will focus on closed-loop active perception, enabling the robotic system to autonomously select optimal interaction strategies based on its current uncertainty or reconstruction error. We aim to further exploit the latent space (e.g., the z -space) to track factors such as geometric deformations. We also want to investigate whether a unified model can be trained to generalize across different skin types when appropriately conditioned on their mechanical properties, toward programmable e-Skins with real-time, tunable compliance⁷. Additionally, integrating the latent filter with reinforcement learning or model-based control could lead to perception-aware policies that exploit the inferred physical properties for more dexterous and reliable manipulation. In conclusion, this study strengthens the importance of jointly optimizing sensor design and interaction strategies, highlighting that tactile perception cannot be decoupled from action or embodiment.

Methods

In this section, we describe the controlled palpation strategies used for tactile interaction and the data processing pipeline employed to prepare the observations for analysis and model training.

Palpation interaction

The e-Skin was mounted on the end-effector of a 7-DOF robotic arm. To generate a rich dataset of interactions, we implemented three robotic palpation primitives (Fig. 2c), combining translational and rotational motions to probe both surface and bulk(internal) object properties. *Pressing* consists of a sinusoidal motion normal to the object surface. *Precession* with a rotational movement involving roll (α) and pitch (β), providing contact with multiple surface regions for probing complex geometries with concave and convex features. *Sliding* emulated natural human texture exploration, capturing surface roughness, structural details, and friction. These movements are defined as:

$$\begin{aligned} \text{pressing} : z(t) &= A_z \cos(2\pi\omega_z t + \pi) \\ \text{precession} : \alpha(t) &= A_\alpha \sin(2\pi\omega_\alpha t + \pi/2) \\ \beta(t) &= A_\beta \sin(2\pi\omega_\beta t) \\ \text{sliding} : x(t) &= A_x \sin(2\pi\omega_x t) \\ y(t) &= A_y \sin(2\pi\omega_y t) \\ \gamma(t) &= A_\gamma \sin(2\pi\omega_\gamma t) \end{aligned} \quad (4)$$

where t represents the time, and the amplitude parameters $A_{(\cdot)}$ and frequency parameters $\omega_{(\cdot)}$ are specific to each movement type. The Kinova robotic arm was programmed within a ROS2⁴⁵ environment to execute these interaction trajectories with precise temporal control. Tactile data from the e-Skin and proprioceptive data from the robot (end-effector pose, joint positions, and applied forces) were synchronously recorded using a custom data acquisition pipeline. A lightweight Protocol Buffers (Protobuf)⁴⁶ schema was implemented to efficiently serialize multi-modal sensor streams, with each message containing a high-resolution timestamp. This ensured time-aligned logging of tactile observations and robot proprioception, providing accurate correspondence between sensor events and executed interaction trajectories for subsequent model training and analysis.

Data processing and training

Two transformations were applied to address the challenges of training on long sequences. For the accelerometer data ($7000 \times 4 \times 4 \times 3$), the normalized magnitude was computed and downsampled to 600 Hz, resulting in a ($6000 \times 4 \times 4$) time-series. A log-mel spectrogram⁴⁷ was then calculated using a window size of 800, hop length of 20, 128 FFT points, and 49 mel bins, yielding a ($300 \times 49 \times 4 \times 4$) representation. This was reshaped into ($300 \times 28 \times 28$) time-series sequences, preserving both spatial structure and temporal dynamics.

For the FSR data ($300 \times 16 \times 16 \times 2$), mean filtering was applied, and the data was downsampled to 30 Hz. To enable evaluation of fusion strategies that integrate spatial (force, shear) and spectral (vibration) information, the FSR signals were spatially rescaled to ($300 \times 28 \times 28 \times 2$) using the interpolation function from ref. 48. This consistent dimensionality across the modalities facilitated comparison between early and late-fusion techniques and supported subsequent ablation studies. A graphical overview of the preprocessing pipeline is shown in Fig. 9a.

Twenty-five percent of the dataset was reserved for testing. The *Latent Filter* model was implemented in PyTorch and trained using the Adam optimizer with a learning rate of 10^{-5} . The training process employed an annealing strategy⁴¹ to gradually increase the regularization weight in the ELBO loss, encouraging stable and effective disentanglement in the latent space. Careful hyperparameter tuning was performed for the learning rate, latent dimensionality, and the temperature parameter of the annealing term.

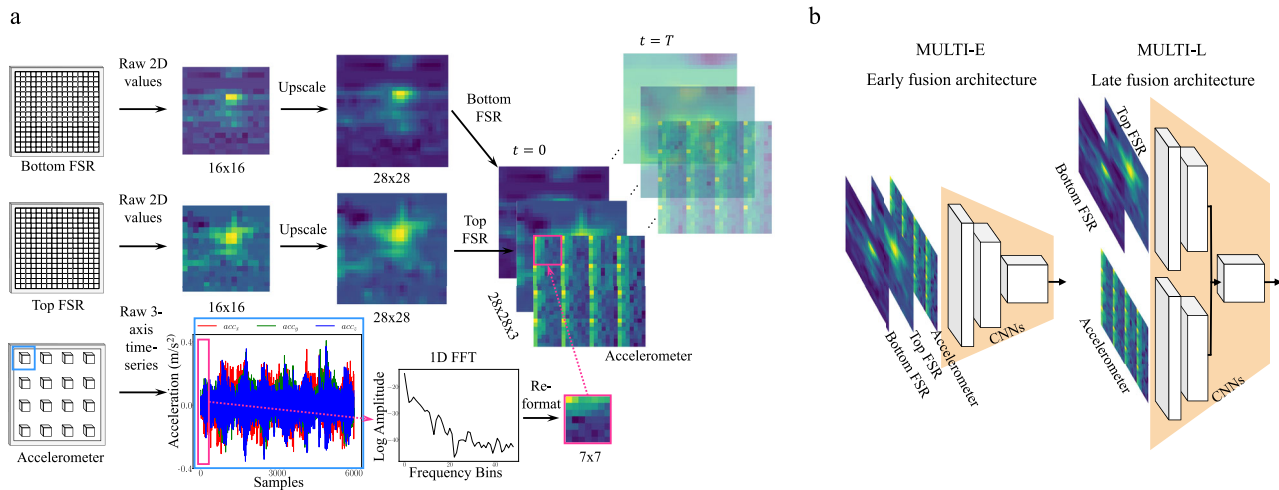


Fig. 9 | Tactile data processing and fusion methods. a Visualization of the pre-processing step of tactile observations. **b** Encoder architecture overview of two fusion techniques to combine accelerometer and FSR observations within the latent filter.

Multi-E represents the early fusion approach, while Multi-L denotes the late-fusion technique.

The latent dimension was $n_s = 32$ (comprising $n_y = 16$ and $n_z = 16$). Model training was performed on a workstation running Ubuntu 20.04 LTS, equipped with an Intel Xeon 5222 CPU, 32 GB RAM, and an NVIDIA RTX A4000 GPU (16 GB VRAM). Each training run converged within approximately 12 h for 500 epochs using a batch size of 48. In total, 24 models - 8 (combination of modality and e-Skin types) $\times$ 3 (interaction primitives) were trained.

Data availability

The full dataset generated and/or analyzed during this study is not publicly available due to its large size (~200 GB), but it is available from the authors upon reasonable request. A representative subset is publicly available at <https://drive.google.com/drive/folders/1vJlzMje5GE4eO1UZ7iM78ToS0qFmxA0?usp=sharing> for evaluation.

Code availability

The code supporting this study is publicly available on the GitHub repository EmbodiedTactileSoftObjects, which can be accessed at <https://github.com/anirvan95/EmbodiedTactileSoftObjects.git>.

Received: 13 August 2025; Accepted: 23 January 2026;

Published online: 12 February 2026

References

- Rawlings, A. E., Bramble, J. P. & Staniland, S. S. Innovation through imitation: biomimetic, bioinspired and biokleptic research. *Soft Matter* **8**, 6675–6679 (2012).
- Johansson, R. S. & Flanagan, J. R. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nat. Rev. Neurosci.* **10**, 345–359 (2009).
- Hayward, V. Is there a plenahaptic function? *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 3115–3122 (2011).
- Hayward, V. Haptic shape cues, invariants, priors and interface design. In *Human Haptic Perception: Basics and Applications* 381–392 (Springer, 2008).
- Zopf, R., Polito, V. & Moore, J. Revisiting the link between body and agency: visual movement congruency enhances intentional binding but is not body-specific. *Sci. Rep.* **8**, 196 (2018).
- Costi, L., Maiolino, P. & Iida, F. How the environment shapes tactile sensing: understanding the relationship between tactile filters and surrounding environment. *Front. Robot. AI* **9**, 930405 (2022).
- Costi, L., Tagliabue, A., Maiolino, P., Clemens, F. & Iida, F. Magneto-active elastomer filter for tactile sensing augmentation through online adaptive stiffening. *IEEE Robot. Autom. Lett.* **7**, 5928–5933 (2022).
- Hughes, J., Scimeca, L., Maiolino, P. & Iida, F. Online morphological adaptation for tactile sensing augmentation. *Front. Robot. AI* **8**, 665030 (2021).
- Devillard, A. et al. Modular, multi-layer e-skin for robotics investigations and applications. In *Proc. IEEE World Haptics Conference (WHC)*, pp. 15–23, (IEEE, 2025).
- Qiu, P. Z., Tan, Y., Thompson, O., Cobley, B. & Nanayakkara, T. Soft tissue characterisation using a novel robotic medical percussion device with acoustic analysis and neural networks. *IEEE Robot. Autom. Lett.* **7**, 11314–11321 (2022).
- Scimeca, L. et al. Action augmentation of tactile perception for soft-body palpation. *Soft Robot.* **9**, 280–292 (2022).
- Herzig, N., Maiolino, P., Iida, F. & Nanayakkara, T. A variable stiffness robotic probe for soft tissue palpation. *IEEE Robot. Autom. Lett.* **3**, 1168–1175 (2018).
- Sun, T., Zhang, Z., Miao, Z. & Zhang, W. A recognition method for soft objects based on the fusion of vision and haptics. *Biomimetics* **8**, 86 (2023).
- Cavdan, M., Doerschner, K. & Drewing, K. Task and material properties interactively affect softness explorations along different dimensions. *IEEE Trans. Haptics* **14**, 603–614 (2021).
- Steer, C., Dinca, T., Jicol, C., Proulx, M. J. & Alexander, J. Feel the force, see the force: exploring visual-tactile associations of deformable surfaces with colours and shapes. In *Proc. CHI Conference on Human Factors in Computing Systems*, 1–13 (Association for Computing Machinery, 2023).
- Visell, Y. & Okamoto, S. Vibrotactile sensation and softness perception. In *Multisensory Softness: Perceived Compliance from Multiple Sources of Information* (ed. Di Luca, M.) 31–47 (Springer, 2014).
- Amin, Y., Gianoglio, C. & Valle, M. Embedded real-time objects' hardness classification for robotic grippers. *Future Gener. Comput. Syst.* **148**, 211–224 (2023).
- Bergmann Tiest, W. M. Tactile perception of material properties. *Vis. Res.* **50**, 2775–2782 (2010).
- Lezkan, A., Metzger, A. & Drewing, K. Active haptic exploration of softness: indentation force is systematically related to prediction, sensation and motivation. *Front. Integr. Neurosci.* **12**, 59 (2018).

20. Xu, Z., Wu, J., Zeng, A., Tenenbaum, J. & Song, S. Densephysnet: learning dense physical object representations via multi-step dynamic interactions. In *Proc. Robotics Science and Systems XV* (Robotics: Science and Systems Foundation, 2019).
21. Scimeca, L., Maiolino, P., Bray, E. & Iida, F. Structuring of tactile sensory information for category formation in robotics palpation. *Auton. Robots* **44**, 1377–1393 (2020).
22. Hui, J. C. & Kuchenbecker, K. J. Evaluating the Biotac's ability to detect and characterize lumps in simulated tissue. In *Proc. Haptics: Neuroscience, Devices, Modeling, and Applications (EuroHaptics)*, Lecture Notes in Computer Science, vol 8619. (Springer, 2014).
23. Konstantinova, J., Cotugno, G., Dasgupta, P., Althoefer, K. & Nanayakkara, T. Palpation force modulation strategies to identify hard regions in soft tissue organs. *PLoS ONE* **12**, e0171706 (2017).
24. Wang, L., Li, Q., Lam, J. & Wang, Z. Tactual recognition of soft objects from deformation cues. *IEEE Robot. Autom. Lett.* **7**, 96–103 (2022).
25. Xu, S., Lin, N., Fan, R., Wu, P. & Chen, X. Exploring hardness and geometry information through active perception. In *Proc. WRC Symposium on Advanced Robotics and Automation (WRC SARA)* 32–37 (IEEE, 2019).
26. Bassal, H., Abbass, Y., Gianoglio, C. & Valle, M. Hardness discrimination using piezoelectric-based biomimetic tactile sensor and machine learning. *IEEE Sens. Lett.* **8**, 1–4 (2024).
27. Caldiran, O., Tan, H. Z. & Basdogan, C. Visuo-haptic discrimination of viscoelastic materials. *IEEE Trans. Haptics* **12**, 438–450 (2019).
28. Huang, J. & Rosendo, A. Variable stiffness object recognition with a CNN-Bayes classifier on a soft gripper. *Soft Robot.* **9**, 1220–1231 (2022).
29. Bednarek, M., Kicki, P., Bednarek, J. & Walas, K. Gaining a sense of touch object stiffness estimation using a soft gripper and neural networks. *Electronics* **10**, 96 (2021).
30. Nonaka, T., Abdulali, A., Sirithunge, C., Gilday, K. & Iida, F. Soft robotic tactile perception of softer objects based on learning of spatiotemporal pressure patterns. In *Proc. IEEE International Conference on Soft Robotics (RoboSoft)* 1–7 (IEEE, 2023).
31. Bicchi, A., Scilingo, E. & De Rossi, D. Haptic discrimination of softness in teleoperation: the role of the contact area spread rate. *IEEE Trans. Robot. Autom.* **16**, 496–504 (2000).
32. Uttayopas, P., Cheng, X., Eden, J. & Burdet, E. Object recognition using mechanical impact, viscoelasticity, and surface friction during interaction. *IEEE Trans. Haptics*, **16**, 251–260 (2023).
33. Estermann, S.-J., Pahr, D. H. & Reisinger, A. Quantifying tactile properties of liver tissue, silicone elastomers, and a 3D printed polymer for manufacturing realistic organ models. *J. Mech. Behav. Biomed. Mater.* **104**, 103630 (2020).
34. Jörmell, H., Devillard, A. & Burdet, E. The identification of haptic invariants in humans and their applications to robotics. *Multisens. Res.* 1(aop), 1–3 (2025).
35. Klushyn, A., Kurle, R., Soelch, M., Cseke, B. & van der Smagt, P. Latent matters: Learning deep state-space models. In *Advances in Neural Information Processing Systems (NeurIPS)* vol. 34, 10234–10245 (Curran Associates, Inc., 2021).
36. Aoyama, M. Y., Moura, J., Saito, N. & Vijayakumar, S. Few-shot learning of force-based motions from demonstration through pre-training of haptic representation. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)* 12839–12845 (IEEE, 2024).
37. Chung, J. et al. A recurrent latent variable model for sequential data. In *Proc. International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15)*, Vol. 2. USA, 2980–2988, (MIT Press, 2015).
38. McInnes, L., Healy, J., Saul, N. & Großberger, L. Umap: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
39. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
40. Davies, D. L. & Bouldin, D. W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* vol. PAMI-1, no. 2, pp. 224–227, (2009).
41. Karl, M., Soelch, M., Bayer, J. & van der Smagt, P. Deep variational bayes filters: Unsupervised learning of state space models from raw data. In *Proc. International Conference on Learning Representations (ICLR)* (OpenReview.net, 2017).
42. Murphy, K. P. Kernel ridge regression. In *Machine Learning: A Probabilistic Perspective* Chap. 14.4.3, 492–493 (MIT Press, 2012).
43. Hillis, J. M., Ernst, M. O., Banks, M. S. & Landy, M. S. Combining sensory information: mandatory fusion within, but not between, senses. *Science* **298**, 1627–1630 (2002).
44. Lederman, S. J. & Klatzky, R. L. Extracting object properties through haptic exploration. *Acta Psychol.* **84**, 29–40 (1993).
45. Macenski, S., Foote, T., Gerkey, B., Lalancette, C. & Woodall, W. Robot Operating System 2: Design, architecture, and uses in the wild. *Sci. Robot.* **7**, eabm6074 (2022).
46. Google Open Source. Protocol buffers: <https://github.com/protocolbuffers/protobuf> Accessed 31 Jul 2025 (2025).
47. McFee, B. et al. librosa: audio and music signal analysis in Python. In *SciPy*, 18–24 (2015).
48. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 8024–8035, (Curran Associates, Inc., 2019).

Acknowledgements

The authors thank Dr. Mohsen Kaboli for discussions on this project. The work was supported in part by the EC H2020 grants INTUITIVE (ITN 861166) and PH-CODING (FETOPEN 829186). Anirvan Dutta was also associated with BMW Group AG, Munich, Germany, during this work.

Author contributions

All authors contributed to the conceptualization, experimental design, and analysis of the research. A. Dutta implemented and developed the latent filtering framework, A. Devillard developed the e-Skin, and Z. Zhang conducted the experimental data collection using robotic hardware. X. Cheng and E. Burdet provided supervision and guidance throughout the project. All authors were involved in the writing and revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44182-026-00077-0>.

Correspondence and requests for materials should be addressed to Etienne Burdet.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026