

Temporally Decoupled Touch-Responsive Humanoid Interaction with Interruptible Motion Execution

Taemoon Jeong^{1†}, Junghyeon Ma^{2†}, Chanwoo Kim¹, Junha Min², Yunseong Bang², Jeongeun Park¹, Sungjoon Choi^{1*}, Kyungseo Park^{2*}

Abstract—A humanoid can be touched while executing an expressive motion, requiring behavior updates without disrupting continuous execution. This is challenging because tactile state, touch-type estimation, and vision-based spatial context operate at different temporal scales. We hypothesize that separating these streams according to their temporal roles enables asynchronous updates while preserving fixed-rate control. We present a touch-responsive humanoid system with a temporally decoupled runtime architecture and interruptible motion execution. A 200 Hz region-wise tactile state supports immediate interruption or hold requests, while slower touch-type and vision-derived occupancy streams update the pending behavior. An anchor-based executor redirects the motion at the next valid anchor. The analytical command-level timing estimate is approximately 14 ms under the assumed stage delays, excluding anchor waiting time and actuator response. A 20-participant study evaluates four touch classes. Across 60 interaction trials, the intended transition was selected in 55 cases (91.7%), and all selected motions satisfied their predefined continuation, hold, or anchor-based redirection outcomes, with no observed robot–human collisions under tested conditions. Project page: <https://tmjeong1103.github.io/TRHI/>

I. INTRODUCTION

Touch conveys attention, affect, and engagement through both contact location and temporal pattern [1], [2]. For a humanoid robot, however, touch-responsive interaction is not only a perception problem. Consider a robot halfway through a greeting when a person grasps its moving arm, taps its torso, or enters the arm’s swept region. The execution system needs to detect the new physical condition, interpret it at the interaction level, and alter the ongoing motion without producing an abrupt joint-space jump.

Existing work provides important parts of this capability. Large-area robot skins localize contact and support reactive or compliant control [3]–[5]. Social-touch classifiers distinguish patterns such as tapping, stroking, and sustained contact [6]–[8]. Vision–tactile systems add intention or proximity context for physical human–robot interaction [9], [10], while socially interactive robots demonstrate touch-centered behaviors such as hugging [11], [12]. These advances do

not directly address how heterogeneous observations should update an expressive motion that is already being executed.

The resulting challenge is temporal as well as functional. Fast tactile monitoring, touch-type estimation, and spatial perception operate at different temporal scales, while motion execution must continue at a fixed command rate. A high tactile sensing frequency alone therefore does not guarantee execution-level responsiveness. A system may detect touch quickly but still wait for slower interpretation or for the current behavior to finish before changing the motion. Replacing trajectories at arbitrary samples can instead introduce discontinuities. Prior HRI studies also show that sensorimotor and communication delays can affect perceived interaction quality [13], [14]. We hypothesize that separating tactile and spatial observations according to their temporal roles enables asynchronous interaction updates while preserving fixed-rate motion execution.

We address this problem through an end-to-end touch-responsive humanoid system. The system integrates a custom 12-DoF upper-body robot, distributed pneumatic and capacitive skin, asynchronous tactile and visual processing, state-based behavior selection, and anchor-based motion execution. A fast region-level tactile state supports interruption or hold requests, while slower touch-type and vision-derived occupancy streams update the pending target behavior. The active trajectory adopts this target at the next validated anchor through a precomputed transition. This temporal separation prevents slower perception from blocking the 200 Hz command pathway while constraining motion changes to validated switching states. We evaluate the architecture through command-level runtime characterization, participant-level cross-validation of touch-type estimation, and integrated physical trials of behavior selection and anchor-based execution. The evaluation distinguishes command-path responsiveness from the delay to the next valid anchor and actuator response.

The contributions are threefold:

- 1) An integrated touch-responsive humanoid system that connects custom robot hardware, distributed multi-modal skin, asynchronous tactile and visual perception, state-based behavior selection, and physical motion execution.
- 2) A temporally decoupled runtime architecture that separates 200 Hz region-level tactile monitoring from slower touch-type estimation and latest-only visual occupancy, allowing asynchronous updates without blocking fixed-rate command execution.

¹Taemoon Jeong, Chanwoo Kim, Jeongeun Park, and Sungjoon Choi are with the Department of Artificial Intelligence, Korea University, Seoul, Republic of Korea (email: taemoon-jeong, chanwoo-kim, baro0906, sungjoon-choi@korea.ac.kr)

²Junghyeon Ma, Junha Min, Yunseong Bang, Kyungseo Park are with the Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Republic of Korea (email: ryan_ma, minion1498, ysbang, kspark@dgist.ac.kr)

[†]Equal contribution.

^{*}Corresponding authors.

- 3) An anchor-based interruptible execution mechanism that combines tactile-change monitoring with asynchronous target-behavior updates and redirects ongoing expressive trajectories through prevalidated transitions.

II. RELATED WORK

A. Robot Skin and Social-Touch Perception

Robot skin research has progressed from localized contact sensing to large-area, distributed, and compliant surfaces for physical interaction [3], [4]. Recent soft and adaptive skins improve conformability, maintainability, and sensitivity over larger robot surfaces [15]–[18]. Whole-body tactile controllers further show that skin signals can support rapid physical reactions and safety-aware behavior [5]. Our hardware builds on this line of work, but our central question is how body-level touch changes an expressive humanoid behavior that is already being executed.

Social-touch perception adds semantic structure to contact. Handcrafted features and learned models have been used to recognize tapping, stroking, pressing, and related patterns [6]–[8]. Semantic recognition alone, however, does not specify when a robot should interrupt an ongoing motion or how to switch without a joint discontinuity. We therefore use a compact classifier in the slower behavior-selection pathway and retain a tactile-state pathway for execution monitoring.

B. Touch-Responsive and Multimodal Robot Interaction

Touch-aware social robots demonstrate value of connecting contact sensing to expressive behavior, particularly in hugging and companion interaction [11], [12]. Systems combine vision and touch to infer intention, attention, or feasible physical response [9], [10]. These studies motivate interpreting touch in embodied context rather than in isolation.

Our work complements these studies by focusing on the runtime organization of direct tactile state, touch-type estimation, spatial context, and ongoing motion execution. These heterogeneous streams are assigned distinct temporal roles so that slower semantic and visual updates do not block tactile monitoring or interruptible motion execution. Vision-based occupancy provides compact spatial conditioning and feasibility filtering for the predefined motion set.

III. TOUCH-RESPONSIVE HUMANOID SYSTEM

A. Platform and Multimodal Skin

The platform is a custom 12-DoF upper-body humanoid with a 2-DoF head and two 5-DoF arms. A head-mounted display presents facial expressions synchronized with upper-body motion. All robot links are 3D-printed to form a lightweight upper-body structure, and joint-position commands are issued through ROS 2 control at 200 Hz.

The skin combines pneumatic pressure pads and capacitive proximity electrodes, as shown in Fig. 1. Each arm contains 32 pneumatic pads. The chest and back each contain four multimodal regions, and the head contains two capacitive regions. After per-pad calibration, the pneumatic channels measure pressures over a range of approximately 0–34.4 kPa. Capacitive sensing detects conductive objects from direct

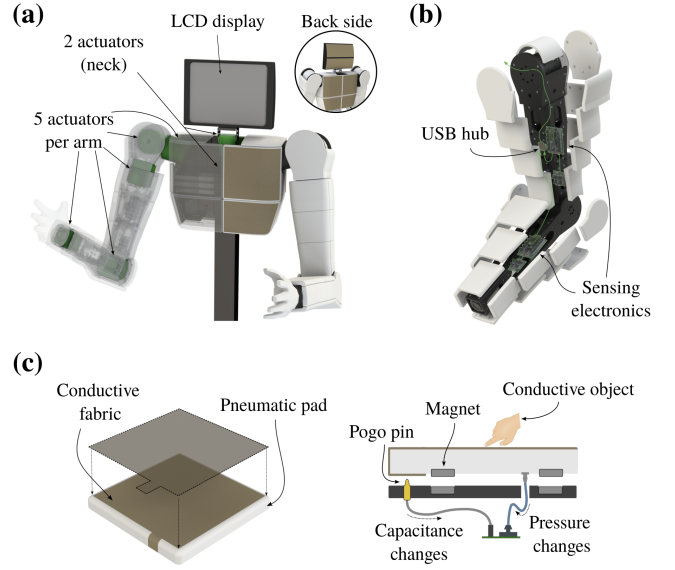


Fig. 1. **Skin-integrated upper-body humanoid.** (a) The 12-DoF platform and instrumented interaction regions. (b) Link-level sensing electronics for local acquisition and aggregation. (c) A multimodal element combining pneumatic pressure and capacitive proximity/contact sensing.

contact to approximately 1–5 cm and complements pressure sensing for light contact and approach.

Local Teensy 4.0 microcontrollers condition and aggregate signals near the robot links. Shared tactile preprocessing produces a body-wide 200 Hz region-wise binary tactile state $\mathbf{b}_t$. A torso-only pathway processes the four-pad chest and back regions separately at 100 Hz, aggregates 2 s of context, and publishes the estimated touch type $\hat{c}_t$ every 0.25 s (4 Hz). Vision updates a latest-only spatial mask $\mathbf{v}_t$ at approximately 30 Hz. Fig. 2 summarizes these temporally decoupled streams and their connections to behavior selection and motion execution.

B. Temporally Decoupled Runtime Architecture

The system uses three computers connected via Gigabit Ethernet: a tactile-processing computer, an interaction computer for state management and spatial occupancy, and a control computer that publishes motor commands at 200 Hz. Most inter-computer communication uses ROS 2/DDS with Best Effort QoS. Each process exposes its latest state through a non-blocking interface, allowing the controller to operate without waiting for slower perception updates. The logical architecture is independent of the number of computers.

Shared tactile preprocessing performs low- and high-pass filtering, thresholding, and region indexing. The resulting state $\mathbf{b}_t$ forms the fast tactile stream and is consumed at 200 Hz by both the behavior selector and the reflex event monitor. It supports direct contact monitoring and higher-priority interruption or hold requests. In parallel, the torso pathway forms the semantic tactile stream and publishes $\hat{c}_t$ at 4 Hz for behavior selection. Human pose observations form the spatial stream and update $\mathbf{v}_t$ through a latest-only buffer for state conditioning and feasible-motion filtering.

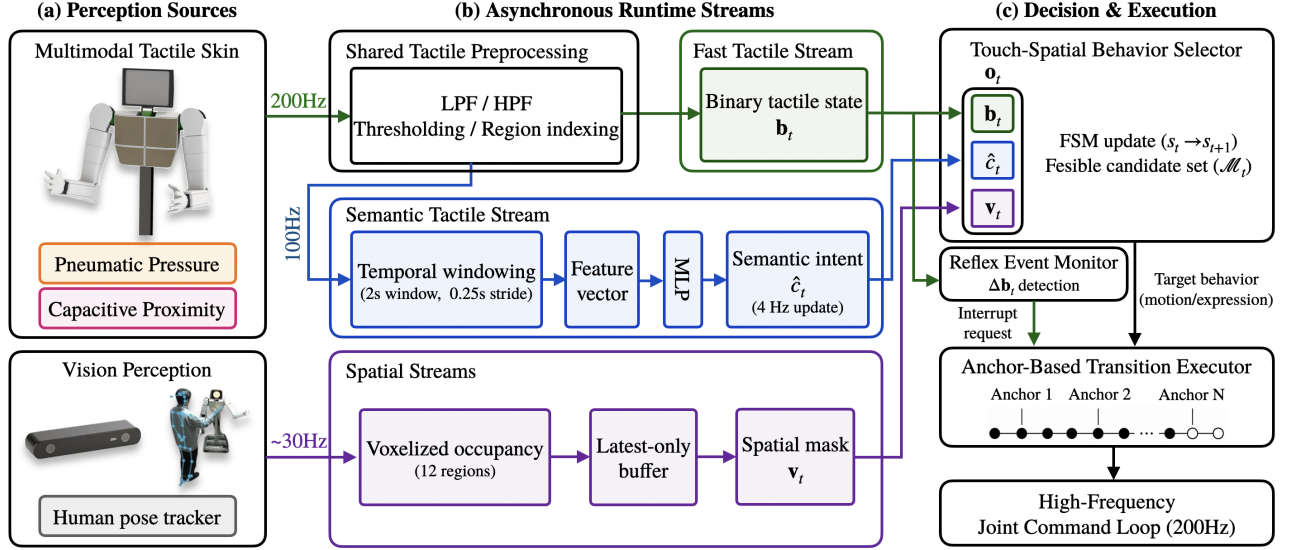


Fig. 2. **Temporally decoupled runtime architecture of the touch-responsive humanoid system.** (a) Multimodal tactile skin and vision provide region-indexed tactile signals and human pose observations. (b) Shared tactile preprocessing produces the 200 Hz tactile state $\mathbf{b}_t$ for direct contact monitoring. The torso touch-type pathway uses a 100 Hz input, a 2 s window, and a 0.25 s stride to publish $\hat{c}_t$ at 4 Hz, while vision updates the latest-only spatial mask $\mathbf{v}_t$ at approximately 30 Hz. (c) The selector uses $\mathbf{o}_t = (\mathbf{b}_t, \hat{c}_t, \mathbf{v}_t)$ to update the finite-state machine (FSM) and feasible candidate set. A reflex monitor detects changes in $\mathbf{b}_t$ and issues a higher-priority interruption request, whereas semantic and spatial updates revise the target behavior asynchronously. The anchor-based executor redirects motion at a valid anchor and continuously publishes joint commands at 200 Hz.

The selector and executor operate as asynchronous producer-consumer processes. The selector uses the newest observation $\mathbf{o}_t = (\mathbf{b}_t, \hat{c}_t, \mathbf{v}_t)$ to update the interaction state and feasible candidate set, then writes the latest target behavior without controlling trajectory timing. Independently, the reflex monitor detects changes between consecutive tactile states and can latch a higher-priority interruption request. The executor publishes joint commands from the active trajectory at 200 Hz and applies a target change only through a validated transition at an allowable anchor. Thus, slower semantic tactile and spatial updates can revise the pending behavior without blocking execution, while direct tactile changes retain priority at the execution layer.

IV. INTERACTION REPRESENTATION AND INTERRUPTIBLE EXECUTION

A. Fast Tactile Stream

Each pneumatic channel is decomposed into a low-pass component for sustained loading and a high-pass component for pressure changes. The high-pass stream uses a 30 Hz cutoff followed by rectification and smoothing. A capacitive channel provides a proximity or light-contact cue. Region-specific thresholds convert these signals into binary pressure, impulse, and proximity indicators, as shown in Fig. 3.

The concatenated region-wise tactile state is denoted by

$$\mathbf{b}_t \in \{0, 1\}^{N_b} \quad (1)$$

where N_b is the total number of binary tactile entries across all body regions and cue types. Its components follow a fixed ordering over body regions and cue types. The active indices therefore encode both the current tactile condition and its location without requiring a separate contact-region variable.

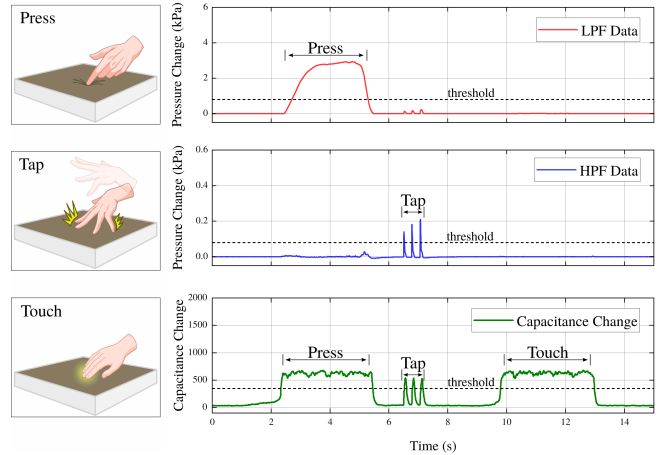


Fig. 3. **Multimodal tactile cues.** Low-pass pressure captures sustained contact, high-pass pressure captures impulsive contact, and capacitance captures proximity or light touch. Dotted lines indicate region-specific binarization thresholds. Here, Press denotes sustained physical pressure applied to the sensor, rather than a distinct semantic touch class.

Because $\mathbf{b}_t$ represents a state rather than an event, the reflex monitor compares consecutive updates:

$$\Delta \mathbf{b}_t = \mathbf{b}_t \oplus \mathbf{b}_{t-1} \quad (2)$$

where $\oplus$ denotes element-wise exclusive OR. Nonzero entries identify changed cues, while the ordered pair $(\mathbf{b}_{t-1}, \mathbf{b}_t)$ distinguishes onset from release and identifies changes in cue type or active body region. A predefined reflex rule forwards a hold or interruption request only when the detected change is relevant to the ongoing motion, without waiting for touch-type interpretation. This fast stream therefore provides direct execution monitoring independently of the slower touch-type pathway.

B. Semantic Tactile Stream

A second pathway estimates the touch class $\hat{c}_t \in \mathcal{C}$, where $\mathcal{C} = \{\text{Idle}, \text{Touch}, \text{Tapping}, \text{Stroke up-down}, \text{Stroke circle}\}$. For the controlled participant protocol, the estimator separately processes the four-pad chest and back regions. The preprocessed cues are sampled at 100 Hz. Each 2 s window is divided into eight 0.25 s bins and advanced with a 0.25 s stride, yielding a 4 Hz touch-type update. Pressure, impulse, and proximity flags from four pads across eight bins form a 96-dimensional binary vector.

An MLP with a 128-unit ReLU hidden layer and five output units maps the vector to $\hat{c}_t$. The MLP is trained for 100 epochs using cross-entropy loss and Adam (learning rate 0.001, batch size 64). The 2 s temporal context supports stroke classification, while new contact conditions remain available through $\mathbf{b}_t$. The touch-type estimate therefore serves as a slower input for behavior selection rather than direct interruption.

The MLP processes only the torso regions. Arm and head interactions are handled through the body-wide tactile state $\mathbf{b}_t$ and region-specific rules in the selector, allowing updates to the pending behavior without torso touch-class estimates.

C. Spatial Stream

Human skeletal joints are tracked using a ZED stereo camera at approximately 30 Hz and used to construct cylindrical and spherical collision geometries in MuJoCo. Intersections between these geometries and 12 predefined workspace regions define the binary occupancy vector

$$\mathbf{v}_t \in \{0, 1\}^{12} \quad (3)$$

Eight outer regions encode coarse approach direction, while four inner regions provide finer spatial resolution near the torso. Fig. 4 shows representative states.

The controller reads $\mathbf{v}_t$ from a latest-only buffer. Occupancy in any workspace region can update the interaction state. Each precomputed motion is labeled with the workspace voxels that it occupies, and motions whose labels overlap the current human occupancy are removed from the candidate set. The controller then selects a behavior from the remaining motions.

These operations remain asynchronous with the tactile streams. The spatial stream therefore provides occupancy-conditioned state updates and feasible-motion filtering without blocking direct tactile monitoring or command execution. It does not provide a general collision-avoidance guarantee.

D. Interaction-Aware Behavior Selection

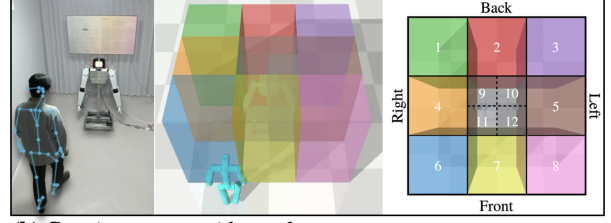
At interaction time, the selector consumes the latest compact observation

$$\mathbf{o}_t = (\mathbf{b}_t, \hat{c}_t, \mathbf{v}_t) \quad (4)$$

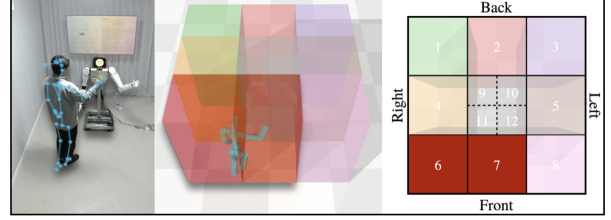
The active indices of $\mathbf{b}_t$ identify the contacted body region and cue type. A finite-state machine (FSM) maintains $s_t \in \mathcal{S}$ and updates it as

$$s_{t+1} = \delta(s_t, \mathbf{b}_t, \hat{c}_t, \mathbf{v}_t) \quad (5)$$

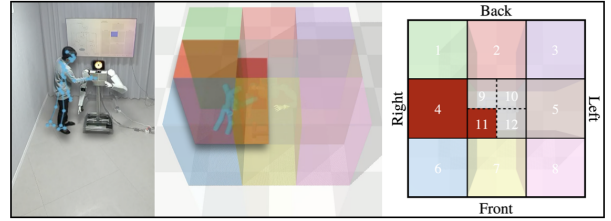
(a) No voxel occupancy during front approach



(b) Greeting gesture with voxel occupancy



(c) Right-front voxel occupancy with partial contact



(d) Right-rear voxel occupancy with partial contact

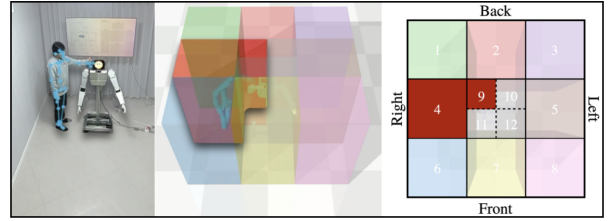


Fig. 4. **Spatial state and feasibility context.** A 12-region occupancy vector changes with frontal, right-front, and right-rear human positions. The representation contributes to FSM updates and constrains candidate upper-body motions. It is not used for identity or detailed pose interpretation.

where δ denotes the predefined interaction-state transition function. The updated state and tactile observation define a candidate motion set.

$$\mathcal{M}_t = \mathcal{M}(s_{t+1}, \mathbf{b}_t, \hat{c}_t) \quad (6)$$

The spatial context is then applied through

$$\mathcal{M}_t^{\text{feas}} = \{m \in \mathcal{M}_t \mid \chi(m, \mathbf{v}_t) = 1\} \quad (7)$$

where $\chi(m, \mathbf{v}_t) = 1$ when the voxel labels assigned to motion m do not overlap the active entries of $\mathbf{v}_t$. Including $\mathbf{v}_t$ in δ allows spatial context to change the interaction state, while χ filters infeasible motions from the resulting candidate set.

An explicit rule table selects the target motion, facial expression, and head orientation from $\mathcal{M}_t^{\text{feas}}$. The selector writes the newest target as the pending behavior but does not determine when the physical trajectory is replaced.

The implementation orders constraints by execution relevance. A spatially infeasible motion is removed from the candidate set and can trigger a new feasible target. A guarded change in $\mathbf{b}_t$ can independently issue a higher-priority hold

or interruption request. Torso touch-type updates select the next interaction behavior among the remaining feasible candidates. This ordering prevents a delayed or transient touch-type estimate from overriding a direct physical condition.

E. Anchor-Based Motion Execution

Each expressive behavior is a precomputed joint-space trajectory, typically 6–8 s long, generated from retargeted human motion and refined through contact-aware motion optimization and learning [19], [20]. The resulting trajectories are checked against the robot’s kinematic limits and annotated with switching anchors. Offline, transition trajectories are generated and validated between compatible anchor states and target-motion initial states. Stochastic path sampling is used to generate the transition set [21], and valid transitions are stored in a lookup table.

During execution, the executor publishes joint commands from the active trajectory at 200 Hz while retaining the newest target behavior from the selector. When the reflex monitor issues an interruption or hold request, the request is latched without waiting for $\hat{c}_t$. Touch-type or spatial updates instead revise the pending target through the selector. Joint-space redirection occurs at the next valid anchor, where the executor retrieves the transition toward the newest feasible target.

Reflex requests take priority over ordinary interaction-state progression, while both follow the same anchor-gated transition rule. Thus, fast response refers to immediate availability of the interruption request in the command pathway, whereas the physical trajectory changes only at a valid anchor.

V. SYSTEM EVALUATION

We evaluate the proposed runtime architecture at three levels. First, runtime characterization examines whether touch-type and vision processing remain isolated from the 200 Hz command loop. Second, semantic tactile validation examines whether the touch-type stream provides a stable input for behavior selection. Third, integrated runtime validation evaluates behavior selection and anchor-based execution under combined tactile and spatial conditions. All experiments use the physical humanoid described in Section III.

A. Runtime Characterization

Table I summarizes an analytical command-level timing estimate based on the implemented update periods and scheduling assumptions. Sensor acquisition and local conditioning contribute up to 5 ms, communication and decision scheduling approximately 4 ms, and alignment to the next 200 Hz command tick up to 5 ms. These terms yield an estimated command-level delay of approximately 14 ms. This estimate does not represent a measured touch-to-command latency distribution or a worst-case bound, and does not account for unquantified timing variability, anchor waiting time, or actuator response.

While touch-type estimation and vision buffering were active, the command process maintained its 200 Hz cadence. Upstream updates changed the pending target without

TABLE I
ANALYTICAL COMMAND-LEVEL TIMING ESTIMATE.

Stage	Delay (ms)
Sensor acquisition and local conditioning	≤ 5.0
Communication and decision scheduling	~ 4.0
Alignment to the next 200 Hz command tick	≤ 5.0
Estimated command-level total	~ 14.0

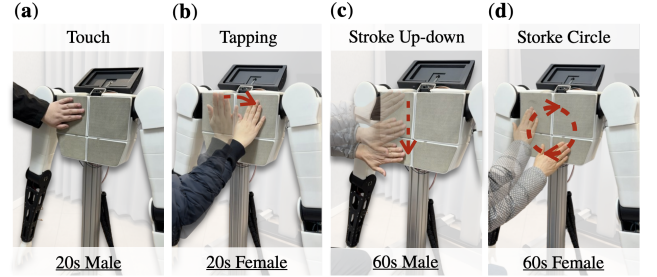


Fig. 5. **Representative tactile gestures used for touch-type estimation.** Participants performed (a) sustained *Touch*, (b) repeated *Tapping*, (c) vertical *Stroke up-down*, and (d) circular *Stroke circle* on the multimodal torso region. Arrow overlays indicate the instructed contact dynamics.

blocking command execution. An interruption request could therefore enter the command pathway through the fast tactile state, while physical trajectory redirection remained gated by the next valid anchor. These observations support temporal isolation between slower perception streams and fixed-rate command execution under the tested conditions.

B. Semantic Tactile Validation

Twenty participants (10 male and 10 female, four per age group from their 20s to 60s) performed four tactile gestures: *Touch*, *Tapping*, *Stroke up-down*, and *Stroke circle*. After brief instructions and a visual demonstration, each participant repeatedly performed the instructed gesture on the designated torso region for approximately three minutes, as illustrated in Fig. 5.¹

Continuous recordings were processed at 100 Hz using the 2 s window and 0.25 s stride described in Section IV. Adjacent windows overlap by 1.75 s (87.5%). The active-class dataset was balanced to 1,000 windows per class and evaluated using five-fold participant-level cross-validation. All windows from each participant were assigned to a single fold, ensuring no participant overlap between training and test sets. The folds were balanced as closely as possible across gender and age groups. Predictions from the held-out folds were aggregated to compute the reported confusion matrices. The online classifier additionally includes an *Idle* state, whereas the confusion matrices report the four instructed active classes.

The mean diagonal of each row-normalized matrix gives macro recall over the four active classes. Based on the aggregated held-out predictions, the corresponding values are approximately 93.8% for participants in their 20s–30s,

¹The study was approved by the Institutional Review Board of Korea University under proposal KUIRB-2025-0192-01.

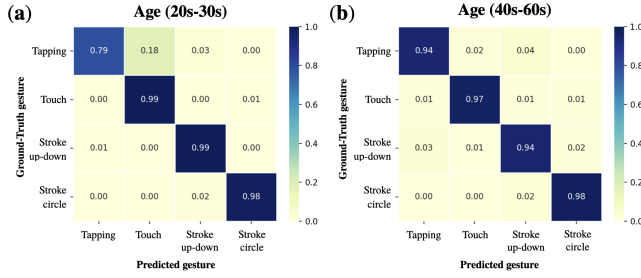


Fig. 6. **Row-normalized touch-type confusion matrices grouped by age.** (a) Participants in their 20s–30s and (b) participants in their 40s–60s. Rows are instructed gestures and columns are predicted gestures.

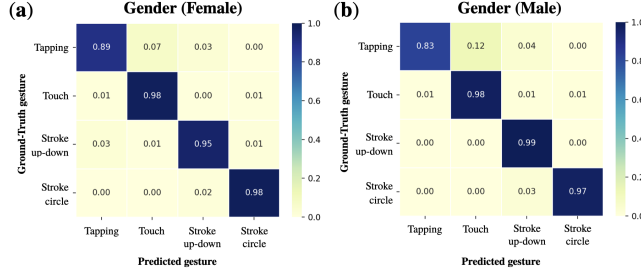


Fig. 7. **Row-normalized touch-type confusion matrices grouped by gender.** (a) Female and (b) male participants.

95.8% for participants in their 40s–60s, 95.0% for female participants, and 94.3% for male participants. The age- and gender-group matrices in Figs. 6 and 7 are descriptive partitions of the same aggregated held-out predictions rather than independent validation results.

The dominant error was *Tapping*→*Touch*, occurring at 18%, 2%, 7%, and 12% in the four reported partitions. The reverse confusion remained between 0% and 1%. This asymmetric error supports retaining the direct impulse/contact pathway independently of the slower touch-type estimate. The results therefore support the semantic stream as a behavior-selection input while preserving the need for a separate fast tactile pathway.

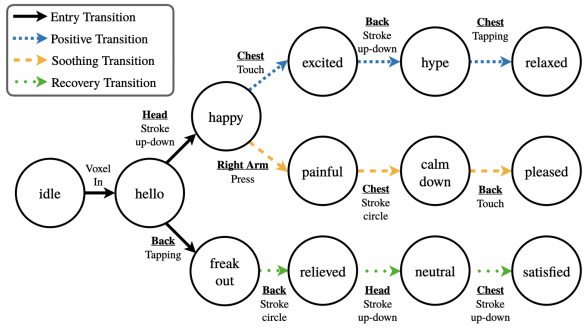
C. Integrated Runtime Validation

Two controller templates, social-affective interaction and attention regulation, were evaluated across six transition branches, as illustrated in Fig. 8. This integrated-runtime evaluation used four participants in a protocol separate from the 20-participant touch-classification evaluation. Each branch was evaluated in 10 trials in total across the four participants, resulting in 60 physical trials. The evaluation examines whether the intended behavior was selected and whether the selected motion was executed according to the predefined tactile and spatial conditions.

A behavior-selection trial was successful when the system recognized the intended contact condition and entered the corresponding predefined FSM state. A failure occurred when the contact was interpreted differently and the intended transition was not selected. As shown in Table II, the FSM selected the intended transition in 55 of 60 trials, corresponding to 91.7%.

All five behavior-selection failures occurred in branches

(a) Social-Affective Interaction Scenario



(b) Attention and Interaction Regulation Scenario

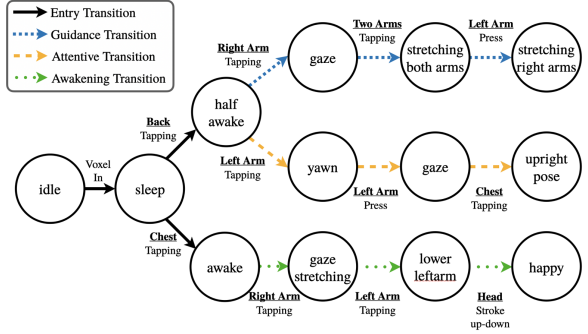


Fig. 8. **Representative interaction scenarios used for integrated runtime validation.** (a) Social-Affective Interaction controller. (b) Attention and Interaction Regulation controller. Each scenario defines predefined interaction states and tactile-triggered transition branches following voxel-based entry into the interaction state.

TABLE II

FSM TRANSITION RELIABILITY ACROSS SIX BRANCHES.

Branch	Success	Trials	Rate
Positive	8	10	80.0%
Soothing	9	10	90.0%
Recovery	8	10	80.0%
Guidance	10	10	100.0%
Attentive	10	10	100.0%
Awakening	10	10	100.0%
Overall	55	60	91.7%

that relied on torso touch-type classification. The MLP-dependent branches achieved 25/30 successful transitions (83.3%), whereas the three branches primarily triggered by direct region-level tactile events achieved 30/30 (100%). The positive and recovery branches each produced two failures, and the soothing branch produced one. The guidance, attentive, and awakening branches bypass the torso MLP and were correctly distinguished through the active region of $\mathbf{b}_t$ in all tested trials.

The same 60 trials were then evaluated for motion-execution reliability after the FSM selected a target motion. A motion-execution trial was successful when the selected motion completed without an observed robot–human collision and matched the predefined execution outcome for that scenario: continuation, hold, or anchor-based redirection according to the imposed voxel occupancy or contact on the moving body region.

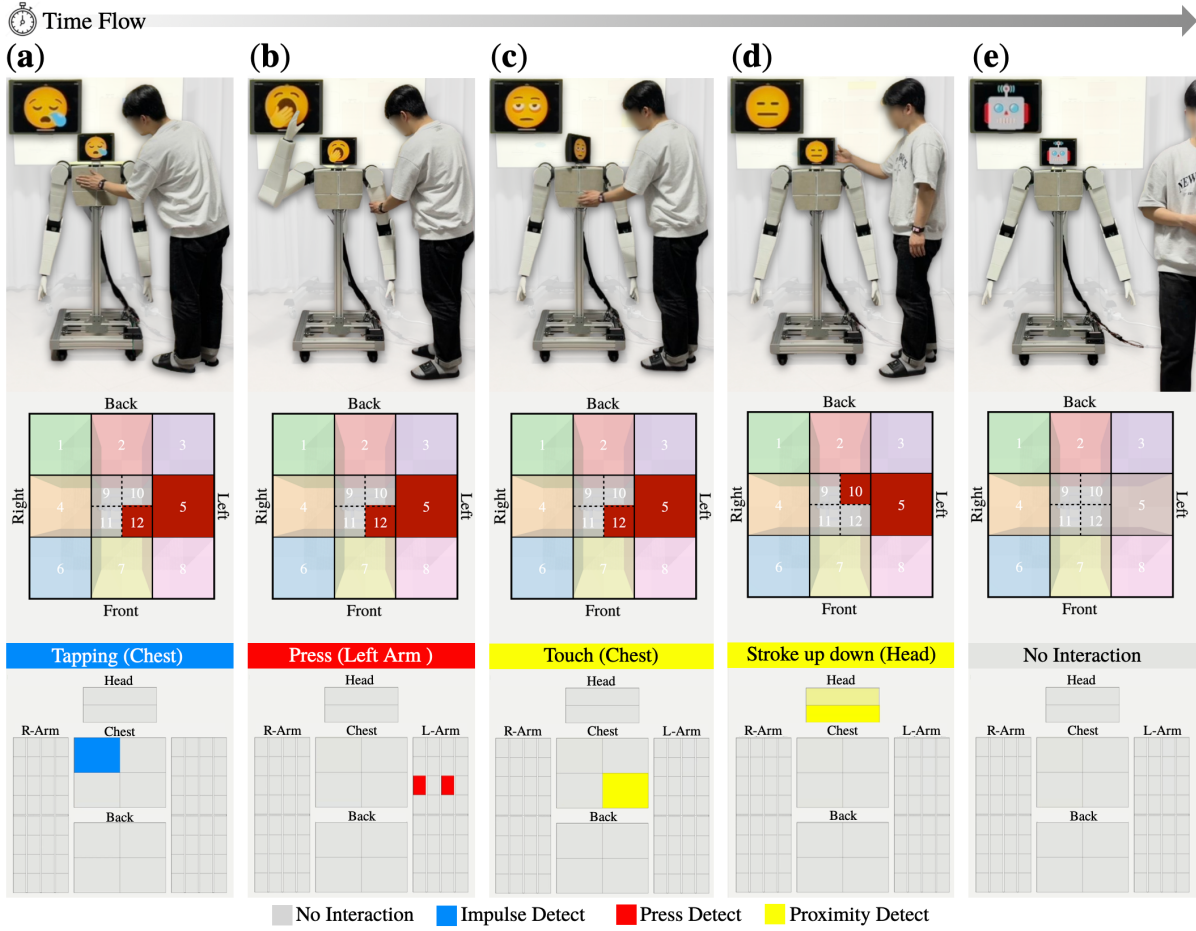


Fig. 9. **Representative touch-responsive execution sequence.** The active tactile cue and robot region change from (a) chest tapping to (b) left-arm pressure, (c) chest touch, (d) head stroking detected through proximity sensing, and (e) no interaction. Facial and upper-body behaviors are redirected through the state controller, while voxel occupancy constrains feasible motion.

All 60 trials (100%) satisfied the predefined execution outcomes for the motions selected by the FSM, with no observed robot–human collisions under the tested conditions. This result reflects motion-execution success after target selection, rather than the correctness of behavior selection, which was 55/60 (91.7%). These results are limited to the predefined motion repertoire and tested occupancy and contact conditions and do not establish general collision avoidance.

Together, these results show that asynchronous target updates, spatial filtering, higher-priority tactile requests, and anchor-based redirection operate within the integrated runtime. Fig. 9 illustrates a representative sequence in which tactile and occupancy updates change the pending behavior while trajectory redirection remains gated by validated anchors. The supplementary video shows the corresponding continuous executions at real-time speed, including tactile interruption, anchor-based redirection, and occupancy-conditioned motion selection.

VI. DISCUSSION AND LIMITATIONS

The main takeaway is the temporal-role separation at the center of the proposed architecture. Direct tactile changes can enter the command pathway without waiting for the

30 Hz spatial stream or the 2 s touch-type context, while the 200 Hz command loop remains isolated from these slower processes. Semantic tactile and spatial updates revise the pending behavior asynchronously, whereas execution-relevant tactile changes can issue a higher-priority interruption or hold request. This priority ordering is predefined rather than universally optimal, but it allows direct tactile requests to reach the executor without arbitrary trajectory replacement. The analytical command-level timing estimate of approximately 14 ms does not include anchor waiting time or actuator response.

The finite motion repertoire and prevalidated transition graph trade online flexibility for predictable execution. Adding new motions requires switching-anchor annotation and offline transition validation. Across 60 physical trials, the FSM selected the intended transition in 55 cases (91.7%), while all 60 trials (100%) satisfied the predefined execution outcomes for the selected motions under the tested conditions. The five selection failures resulted from incorrect torso touch-type estimates. These results characterize behavior selection and motion execution within the predefined repertoire, rather than open-ended behavior generation or transitions to unseen motions.

Several limitations remain. The coarse 12-region oc-

cupancy representation supports state conditioning and feasible-motion filtering but does not capture detailed human geometry, body-part semantics, orientation, or crowded multi-person scenes, nor does it guarantee collision avoidance. The current setup assumes a single human interacting with a single robot, while multi-human settings would require source attribution and arbitration among simultaneous interaction requests. Five-fold participant-level cross-validation evaluated generalization within the controlled protocol, whereas unconstrained touch remains unevaluated. The reported timing estimate does not establish perceived responsiveness, reciprocity, or motion naturalness. Moreover, the absence of a baseline without temporal decoupling limits direct attribution of the observed runtime behavior to the proposed architecture. Future work should examine comparative runtime evaluations and subjective HRI studies of perceived responsiveness, naturalness, and comfort, alongside richer spatial representations, broader touch data, and adaptive skin baseline compensation.

VII. CONCLUSION

We presented a temporally decoupled runtime architecture for touch-responsive humanoid interaction. Fast tactile monitoring remains available for interruption or hold requests, while slower touch-type and vision-based spatial updates revise the pending behavior without blocking the 200 Hz command loop. Ongoing motions are redirected only through validated anchors, separating behavior updates from physical trajectory replacement.

The analytical command-level timing estimate is approximately 14 ms under the stated assumptions, excluding anchor waiting time and actuator response. Across 60 physical trials, the intended FSM transition was selected in 55 cases (91.7%), and all selected motions matched their predefined execution outcomes under the tested tactile and spatial conditions. These results support temporal-role separation as a practical way to connect heterogeneous perception streams with continuous, interruptible humanoid motion execution.

ACKNOWLEDGMENT

This work was supported by grants from the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT): (No. RS-2019-II190079, Artificial Intelligence Graduate School Program (Korea University), 20%, No. RS-2022-II220480, Development of Training and Inference Methods for Goal-Oriented Artificial Intelligence Agents, 20%, and No. RS-2024-00336738, Development of Complex Task Planning Technologies for Autonomous Agents, 20%, RS-2025-25442149, LG AI STAR Talent Development Program for Leading Large-Scale Generative AI Models in the Physical AI Domain, 20%). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT): (No. RS-2026-25547694, Development of Embodiment-Agnostic General

Intelligence and Learning Transfer Technologies for Autonomous Complex-Task Execution Across Heterogeneous Humanoids, 20%)

REFERENCES

- [1] M. J. Hertenstein, D. Keltner, B. App, B. A. Bulleit, and A. R. Jaskolka, "Touch communicates distinct emotions," *Emotion*, vol. 6, no. 3, pp. 528–533, 2006.
- [2] I. Morrison, L. S. Loken, and H. Olausson, "The skin as a social organ," *Experimental Brain Research*, vol. 204, no. 3, pp. 305–314, 2010.
- [3] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing—from humans to humanoids," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 1–20, 2010.
- [4] P. Mittendorf and G. Cheng, "Humanoid multimodal tactile-sensing modules," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 401–410, 2011.
- [5] S. Jiang and L. L. Wong, "A hierarchical framework for robot safety using whole-body tactile sensors," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 8021–8028.
- [6] M. M. Jung, M. Poel, R. Poppe, and D. K. J. Heylen, "Automatic recognition of touch gestures in the corpus of social touch," *Journal on Multimodal User Interfaces*, vol. 11, pp. 81–96, 2017.
- [7] H. Choi, D. Brouwer, M. A. Lin, K. T. Yoshida, C. Rognon, B. Stephens-Fripp, A. M. Okamura, and M. R. Cutkosky, "Deep learning classification of touch gestures using distributed normal and shear force," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 3659–3665.
- [8] L. Chen, Y. Zhu, and M. Li, "Tactile-GAT: Tactile graph attention networks for robot tactile perception classification," *Scientific Reports*, vol. 14, no. 1, p. 27543, 2024.
- [9] C. Y. Wong, L. Vergez, and W. Suleiman, "Vision- and tactile-based continuous multimodal intention and attention recognition for safer physical human–robot interaction," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 3, pp. 3205–3215, 2024.
- [10] M. Iskandar, A. Albu-Schaffer, and A. Dietrich, "Intrinsic sense of touch for intuitive physical human–robot interaction," *Science Robotics*, vol. 9, no. 93, p. eadn4008, 2024.
- [11] A. E. Block, S. Christen, R. Gassert, O. Hilliges, and K. J. Kuchenbecker, "The six hug commandments: Design and evaluation of a human-sized hugging robot with visual and haptic perception," in *Proceedings of the 2021 ACM/IEEE International Conference on Human–Robot Interaction*, 2021, pp. 380–388.
- [12] Q. Ren and T. Belpaeme, "Tactile interaction with social robots influences attitudes and behaviour," *International Journal of Social Robotics*, vol. 16, no. 11, pp. 2297–2317, 2024.
- [13] M. K. X. J. Pan, E. Knoop, M. Bacher, and G. Niemeyer, "Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 6735–6741.
- [14] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, "How quickly should a communication robot respond? delaying strategies and habituation effects," *International Journal of Social Robotics*, vol. 1, no. 2, pp. 141–155, 2009.
- [15] K. Park, H. Yuk, M. Yang, J. Cho, H. Lee, and J. Kim, "A biomimetic elastomeric robot skin using electrical impedance and acoustic tomography for tactile sensing," *Science Robotics*, vol. 7, no. 67, p. eabm7187, 2022.
- [16] K. Park, K. Shin, S. Yamsani, K. Gim, and J. Kim, "Low-cost and easy-to-build soft robotic skin for safe and contact-rich human–robot collaboration," *IEEE Transactions on Robotics*, 2024.
- [17] M. J. Yang, K. Park, W. D. Kim, and J. Kim, "Robotic skin mimicking human skin layer and pacinian corpuscle for social interaction," *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 4, pp. 2709–2719, 2024.
- [18] M. J. Yang, H. Chung, Y. Kim, K. Park, and J. Kim, "A body-scale robotic skin using distributed multimodal sensing modules: Design, evaluation, and application," *IEEE Transactions on Robotics*, vol. 41, pp. 96–109, 2025.
- [19] T. Jeong, T. Byun, J. Kim, K. Choi, J. Oh, S. Lee, O. Darwish, J. Kim, and S. Choi, "Robust and expressive humanoid motion retargeting via optimization-based rig unification," in *2025 IEEE/RSJ International*

Conference on Intelligent Robots and Systems (IROS), 2025, pp. 21 619–21 626.

- [20] T. Jeong, Y. Chai, S. Choi, J. Bak, C. Kim, J. Yoon, Y. Lee, J. Lee, K. Lee, J. Kim, and S. Choi, “CoRe: A hybrid approach of contact-aware optimization and learning for humanoid robot motions,” in *2025 IEEE-RAS 24th International Conference on Humanoid Robots (Humanoids)*, 2025, pp. 293–300.
- [21] S. Choi, K. Lee, and S. Oh, “Gaussian random paths for real-time motion planning,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 1456–1461.