Design and Evaluation of a Dual-Layer Emotion —Personality Framework for Adaptive Conversational Robots

Shitara Kaede [1], Kantawatchr Chaiprabha [2], Pimolkan Piankitrungreang [2], Kosuke Takano [1,*], Ratchatin Chancharoen [2,*], and Gridsada Phanomchoeng [2,*]

[1] Department of Information and Computer Sciences, Faculty of Information Technology, Kanagawa Institute of Technology, Kanagawa, Tokyo, Japan

[2] Department of Mechanical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand

Email: s2585011@cco.kanagawa-it.ac.jp (S.K.); kantawatchr@gmail.com (K.C.); 6770266021@student.chula.ac.th (P.P); takano@ic.kanagawa-it.ac.jp (K.T.); ratchatin.C@chula.ac.th (R.C.); gridsada.phanomchoeng@gmail.com (G.P.)

*Corresponding author

***Abstract*—This paper presents a dual-layer emotional framework for human–robot conversational interaction that integrates internal emotion, representing the robot's intrinsic affective state, and social emotion, representing outward emotional expression adapted for interpersonal alignment. Unlike conventional dialogue systems that rely primarily on semantic and contextual interpretation, the proposed framework processes user input through three complementary dimensions: content, context, and emotion, while supporting multimodal interaction through voice and physical actions. Emotion computation is regulated using two personality-modulated parameters. Sensitivity controls how strongly external stimuli influence internal emotional states, whereas consideration governs the degree to which the robot aligns its social expression with the user's affect. A robotic prototype was developed to implement the framework, integrating multimodal sensors and expressive actuators for interactive operation. Preliminary experiments were conducted to evaluate the temporal evolution of emotional states, system response latency, and the influence of personality parameters on emotional behavior. The results illustrate that the system can update internal and social emotional states dynamically, adapt responses according to personality parameters, and generate emotionally coherent multimodal outputs. From a Human–Computer Interaction (HCI) perspective, the proposed framework provides a system-level approach for designing conversational interfaces capable of emotionally adaptive multimodal interaction in embodied robotic systems. The findings suggest that personality-modulated emotional regulation can support more flexible and context-dependent conversational behavior compared with conventional reactive emotional dialogue mechanisms.**

***Keywords*—conversational AI, emotional intelligence, social robotics, personality modulation, human–robot interaction**

I. INTRODUCTION

Recent advances in conversational AI, driven by large-scale pretraining and transformer-based dialogue models, have significantly improved the fluency and contextual relevance of open-domain dialogue systems. While representative work such as DialoGPT and recent foundation models demonstrate strong linguistic capability, practical deployment increasingly reveals limitations in sustaining engaging interaction, as user satisfaction depends not only on language quality but also on socially coherent behavior [1−3]. These trends motivate research that moves beyond content generation toward conversational systems capable of managing interaction quality at a social and affective level.

In embodied Human–Robot Interaction (HRI), affect-aware behavior becomes particularly important because physical presence amplifies user expectations of responsiveness and social intent. Prior studies in social robotics and multimodal affective computing consistently report that emotion understanding and expression benefit from integrating multiple interaction channels, including speech and physical cues such as touch [4−6]. Moreover, personality-related factors have been shown to influence robot acceptance and interaction outcomes, suggesting the need for mechanisms that regulate how emotion is internally maintained and socially expressed over time [7, 8]. Together, these findings motivate engineering frameworks for conversational robots that integrate multimodal affect cues and personality-modulated emotional dynamics within real robotic systems.

Despite substantial progress in affective computing and social robotics, many conversational robot systems remain limited by reactive emotion handling, where user affect is detected through sentiment or emotion recognition and directly mapped to predefined responses [9, 10]. Such approaches typically lack explicit modeling of internal affective dynamics, resulting in transient emotional

Manuscript received February 9, 2026; revised March 10, 2026; accepted April 3, 2026; published June 22, 2026.

behavior and limited consistency across interactions, a limitation repeatedly noted in surveys of social and affective robots [11]. Although personality has been shown to influence human–robot interaction outcomes, it is often treated as a static design attribute or evaluated at the perceptual level, rather than implemented as a tunable computational mechanism within the robot's emotional processing architecture [8, 12]. Furthermore, many affect-aware conversational systems focus on linguistic interaction alone, without integrating physical modalities such as touch within a unified emotional framework [13, 14]. These limitations reveal a gap between conceptual models of emotional interaction and engineering implementations capable of integrating multimodal sensing, internal emotion evolution, personality-modulated adaptation, and system-level evaluation in embodied conversational robots.

Recent advances in multimodal affective and sentiment analysis suggest that emotion understanding in interactive systems can benefit from integrating multiple information channels rather than relying on text alone. Transformer-based and attention-driven fusion architectures have enabled effective modeling of unaligned multimodal sequences, improving robustness under noisy or incomplete inputs [15]. Several post-2019 surveys consolidate progress in multimodal sentiment analysis, summarizing fusion strategies, benchmark datasets, and remaining challenges such as modality imbalance and cross-domain generalization [9, 10, 16–18]. In parallel, conversational AI has advanced rapidly through large-scale dialogue pretraining and instruction-following paradigms. Generative dialogue models such as DialoGPT and subsequent instruction-tuned language models have substantially improved conversational coherence and controllability [19, 20], while recent foundation models extend conversational capability to multimodal inputs [21]. Despite these advances, emotion in most conversational systems remains treated as an auxiliary signal rather than a persistent internal state influencing response generation.

Within human–robot interaction, social robotics research consistently emphasizes that affective and social cues play an important role in user acceptance and sustained engagement, particularly for embodied conversational robots [22]. Surveys on social touch highlight physical interaction as a meaningful affective communication channel that remains underexplored in system-level conversational frameworks [23]. Furthermore, extensive meta-analytic and systematic review studies show that both human personality traits and robot personality design significantly influence interaction outcomes, acceptance, and trust [24–26]. However, existing approaches often model personality conceptually or perceptually, rather than embedding it as a tunable computational parameter within the robot's affective architecture [27, 28]. In addition, reviews of long-term HRI indicate that many evaluations remain short-term, limiting insight into emotional consistency and adaptation across repeated interactions [29]. These limitations motivate engineering-oriented frameworks that integrate multimodal affect cues, internal emotion dynamics, personality-modulated adaptation, and real robotic implementation.

This paper presents an engineering framework for emotionally adaptive conversational robots. The study focuses on modeling how emotion is internally maintained and socially expressed during interaction, and on regulating this process through personality-inspired parameters. The proposed approach is designed for implementation in embodied robotic systems and supports near real-time interaction using multimodal inputs.

The main contributions of this work are threefold.

First, the paper proposes a dual-layer emotional framework that explicitly separates internal emotion, representing the robot's affective state evolution, from social emotion, governing outward emotional expression during conversation. This separation enables emotional continuity while allowing context-dependent adaptation.

Second, a personality-modulated emotion computation mechanism is introduced, in which tunable parameters control how external stimuli influence internal emotion and how internal emotion is translated into social behavior. This provides a computationally efficient means to generate diverse interaction styles without modifying the system architecture.

Third, the proposed framework is implemented on an embodied conversational robot integrating speech and tactile interaction, and its feasibility is demonstrated through system-level evaluation showing adaptive and emotionally coherent behavior during interaction.

The remainder of this paper is organized as follows. Section II introduces the proposed dual-layer emotional interaction framework and its conceptual foundations. Section III describes the system architecture and personality-modulated emotional processing mechanisms. Section IV presents the fuzzy emotion modeling and computational formulation. Section V details the hardware prototype implementation. Section VI explains the experimental setup and evaluation methodology, followed by the experimental results in Section VII. Section VIII discusses the findings and limitations, and Section IX concludes the paper with directions for future work.

II. Proposed Emotional Interaction Framework

The proposed emotional interaction framework establishes a conceptual model for integrating emotion and personality into conversational human–robot interaction. As illustrated in Fig. 1, communication between a human and a robotic agent is modeled as a bidirectional process consisting of two sequential interaction stages: message transmission from the human to the robot and response generation from the robot to the human. These interaction stages describe communication flow and are conceptually separated from the internal emotional computation layers implemented within the robotic system.

To characterize communication signals, each interaction stage is analyzed through three complementary dimensions: context, content, and emotion. The context dimension represents situational and interaction-related conditions, including conversational history and environment. The content dimension describes semantic information conveyed through communication. The emotion dimension represents affective cues expressed through linguistic, paralinguistic, or physical interaction channels.

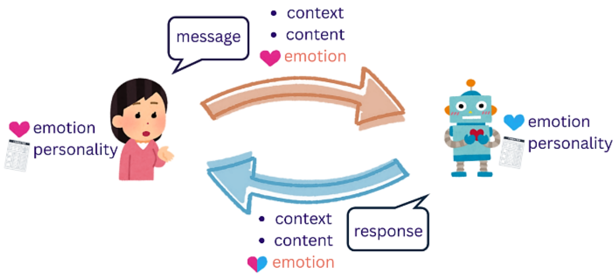


Fig. 1. Conceptual overview.

The framework is motivated by two behavioral observations in human social interaction. First, emotional states evolve dynamically through interpersonal communication. Second, individuals tend to adapt their behavior to align with the emotional states of their interaction partners. These observations highlight the need for conversational robotic systems that maintain emotional continuity while generating socially appropriate responses.

Based on these observations, the proposed framework introduces a dual-layer emotional representation that separates internal emotional state evolution from outward emotional expression. This separation enables robotic agents to preserve consistent internal emotional dynamics while adapting external emotional behavior to social context. The computational realization of this architecture is described in the following section.

III. Emotional Processing and Personality-Modulated Architecture

This section describes the computational architecture that governs affective perception, emotional state evolution, and response generation in the proposed conversational robotic system. The architecture integrates multimodal input processing, dual-layer emotional representation, and personality-modulated adaptation to support emotionally coherent human–robot interaction.

A. Overall System Architecture and Conversation Flow

The overall system architecture is illustrated in Fig. 2. The proposed framework supports multimodal input, layered emotional computation, and adaptive response generation within a unified conversational pipeline.

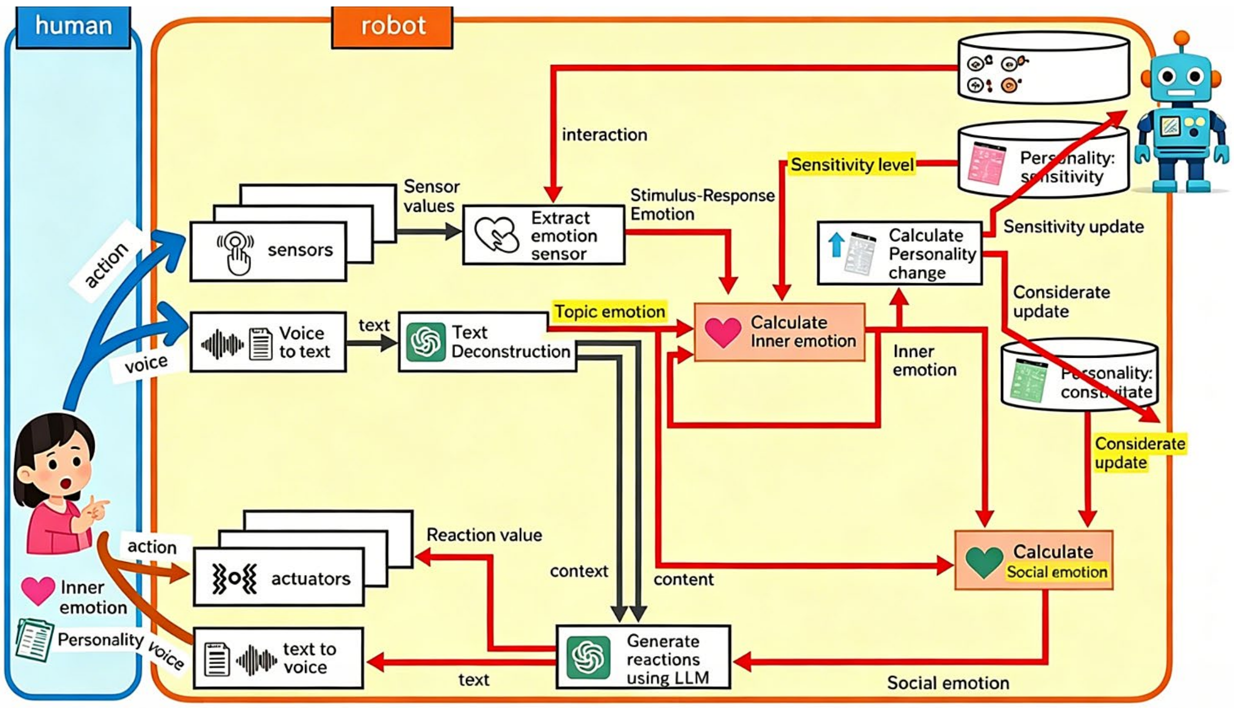


Fig. 2. Conversation flow diagram.

User input is received through two primary channels: voice and physical interaction. Voice signals are first converted into text and processed by a text analytics module that extracts semantic content, contextual meaning, and topic-level emotional cues. In parallel, physical user actions such as touch or gesture are detected through sensors, where each sensor input is mapped to a predefined emotional stimulus stored in an emotional mapping table. These multimodal inputs collectively form the basis for affective perception.

The perceived emotional information is processed through the dual-layer emotional architecture, which maintains internal emotional continuity while regulating outward emotional expression. Emotional states are stored in an emotional state table, where each emotion is represented as a quantitative value associated with predefined emotion labels. At each interaction step, the dominant emotion is selected based on the highest activation level.

Contextual information, semantic content, internal emotional state, and social emotional expression are subsequently provided as conditioning inputs to a Large Language Model (LLM). The LLM generates responses that reflect both conversational meaning and emotional state. The generated output is then rendered through text-to-speech synthesis and physical actuator control, enabling coordinated verbal and non-verbal emotional expression.

B. *Dual-Layer Emotional Representation*

The proposed system models robotic affect using two interrelated emotional layers: internal emotion and social emotion.

The internal emotion layer represents the robot's intrinsic affective state, which evolves dynamically over time based on interaction history, perceived user emotion, and multimodal sensory input. Maintaining an internal emotional state enables temporal continuity and emotional memory across repeated interactions.

The social emotion layer governs outward emotional expression during interaction. This layer regulates how internal emotional states are translated into observable behaviors, including verbal responses, speech prosody, and physical interaction cues. The separation between internal and social emotional layers enables the robotic agent to maintain emotionally consistent internal states while adapting external behavior to conversational context.

C. *Multimodal Emotional Perception*

The emotional architecture incorporates multimodal perception to estimate user affective state. Communication inputs are decomposed into context, content, and emotion dimensions. Emotional cues are extracted from linguistic signals, speech characteristics, and physical interaction channels.

Multimodal integration improves emotional recognition robustness by reducing reliance on individual modalities and supporting reliable interpretation under uncertain or incomplete input conditions.

D. *Personality-Modulated Emotional Adaptation*

To regulate emotional adaptation, the architecture incorporates tunable personality parameters that influence emotional state evolution and response generation.

The sensitivity parameter determines the degree to which external stimuli influence the robot's internal emotional state. Higher sensitivity produces stronger responsiveness to user emotional input, whereas lower sensitivity promotes emotional stability.

The consideration parameter controls the relative weighting between user emotional input and the robot's internal emotional state during response generation. Higher consideration emphasizes empathetic alignment with user affect, while lower consideration prioritizes internal emotional consistency.

These parameters enable the robotic agent to generate diverse conversational styles while preserving a unified computational framework.

E. *Emotional State Transition and Response Generation*

The emotional processing pipeline operates in three stages. First, multimodal inputs are interpreted to estimate user emotional state. Second, the internal emotion layer updates the robot's intrinsic affective state using personality-modulated adaptation rules. Third, the social emotion layer generates outward emotional expressions that guide conversational response selection and behavioral output.

This layered processing structure enables emotionally coherent and socially adaptive interaction suitable for near real-time embodied conversational systems.

IV. Fuzzy Emotion Modeling and Computational Formulation

This section presents the mathematical formulation of the proposed emotion processing framework. Fuzzy logic is employed to model, update, and express the robot's affective states by addressing perceptual uncertainty and integrating multimodal sensory inputs with personality-driven modulation parameters. The proposed formulation provides a computational interpretation of affect regulation, enabling emotionally coherent and socially adaptive human–robot interaction.

A. *Fuzzy Mapping of Sensory Stimuli*

The definitions of symbols and variables used in the fuzzy emotion model are summarized in Table I. The proposed approach integrates perceptual uncertainty, multimodal sensory inputs, and personality-modulated parameters to support emotionally coherent and socially adaptive interaction. Through this formulation, the framework provides a computational representation of affect modulation that is consistent with established theories of emotional regulation and social adaptation.

TABLE I. Definition of Symbols and Variables Used in the Fuzzy Emotion Model

Symbol / Variable	Description
i, d	Tactile intensity and duration
E	Set of basic emotions {happy, sad, angry}
x_s, x_i, x_c, x_{soc}	The membership vectors corresponding to stimulus, inner, context, and social emotions.
T	Set of tactile categories {No touch, Tap, Press, Stroke, Push}
m	A tactile membership degree vector
F_{expr}	Expressed face, Categorical face output
α	Dominance emotion of social emotion
λ	Intensity of social emotion
$w_{stim}, w_{cont}, w_{sens}, w_{cons}$	Weight effects for stimulus, context, sensitivity, and considerate emotions.

To ensure reproducibility and interpretability, the fuzzy membership functions were designed to represent the gradual and uncertain nature of affective interaction. Emotional cues derived from tactile interaction and conversational context are inherently continuous rather than discrete. Therefore, fuzzy membership functions were adopted to enable smooth transitions between emotional states and to avoid abrupt switching near decision boundaries. Overlapping membership functions were intentionally used to maintain emotional continuity and stable behavioral responses during interaction.

In addition, the personality parameters, namely sensitivity and consideration, were defined within a normalized range of [0,1] to provide an interpretable and reproducible control mechanism. The sensitivity parameter regulates how strongly external stimuli influence the internal emotional state, while the consideration parameter controls the balance between internal emotion and socially expressed emotion. Representative values of 0, 0.5, and 1 were used in the illustrative experiments to demonstrate low, balanced, and high responsiveness of the emotional framework. This design facilitates reproducibility and allows consistent implementation of the proposed emotional interaction model.

The emotion generation process begins with the fuzzy mapping of tactile inputs into stimulus emotions. Each tactile input channel, denoted as τ_k, is evaluated using a corresponding fuzzy membership function $\mu_{\tau_k}(i,d)$, where i and d represent the input intensity and duration, respectively:

$$m_k = \mu_{\tau_k}(i,d) \tag{1}$$

The overall tactile emotional activation is represented as a membership vector:

$$m = \left[m_1, m_2, \ldots, m_{|T|}\right]^{\mathsf{T}} \tag{2}$$

A linear mapping matrix R is then used to project the fuzzy activation vector into the emotional stimulus space:

$$x_s = \sum_{k=1}^{|T|} m_k \times r_k = R \times m \tag{3}$$

This mapping formalizes how different tactile interaction patterns contribute to stimulus-driven emotional responses. It enables multimodal sensory inputs to be integrated and weighted according to predefined emotional associations.

B. Inner Emotion State Update

The robot's internal emotion, denoted as x_i, represents its intrinsic affective state. This state evolves dynamically over time as a function of prior emotional state, sensory stimuli, and conversational context. The temporal evolution of the internal emotion is modeled using the following update equation:

$$x_i(t+1) = (1 - w_{sens}) \times x_i(t) + w_{sens} \times \frac{w_{stim} \times x_s(t) + w_{cont} \times x_c(t)}{w_{stim} + w_{cont}} \tag{4}$$

The parameter w_{sens} represents the sensitivity level, which determines the degree to which external stimuli influence the internal emotional state. The parameters w_{stim} and w_{cont} denote the relative contributions of stimulus-driven emotion and context-driven emotion, respectively.

This formulation enables gradual emotional state transitions while allowing personality-modulated responsiveness to external interaction cues, thereby supporting stable yet adaptive emotional behavior.

C. Social Emotion Computation

The social emotion, denoted as x_{soc}, represents the robot's outward emotional expression during interaction. This layer reflects a combination of the robot's internal affective state and its adaptation to the emotional state of the interaction partner. The social emotion is computed as follows:

$$x_{soc} = (1 - w_{cons}) \times x_c + w_{cons} \times x_i \tag{5}$$

where w_{cons} denotes the considerate level, a personality-modulated parameter that regulates the balance between user-oriented and self-oriented emotional behavior.

A higher considerate level increases the influence of the robot's internal emotional state on outward emotional expression, promoting emotionally consistent behavior. Conversely, a lower considerate level strengthens alignment with the user's perceived emotional state, enabling adaptive and empathetic interaction.

This formulation enables flexible regulation of emotional expression, allowing the robotic agent to dynamically adjust its social behavior while preserving emotional coherence across interactions.

D. Dominant Emotion and Intensity Estimation

After computing the social emotion vector, the dominant emotional state is identified as the component with the highest activation level. The dominant emotion label is determined by:

$$\alpha = arg\max_j \left(x_{soc,j}\right) \tag{6}$$

To quantify the intensity of the expressed emotion, the contrast between the strongest and second strongest emotional components is computed. First, the index of the dominant emotional component is defined as:

$$j^* = \arg\max_j x_{soc,j} \tag{7}$$

The second strongest emotional activation is then obtained by:

$$x_{second} = \max_{j \neq j^*} x_{soc,j} \tag{8}$$

The emotional intensity is defined as the difference between the dominant and second strongest emotional components:

$$\lambda = x_{soc,j^*} - x_{second} \tag{9}$$

The parameter λ represents the clarity and strength of the expressed emotion. A larger value of λ indicates a more dominant and clearly distinguishable emotional state, which influences the expressiveness and behavioral vividness of the robot during interaction.

E. Emotional Expression Mapping

The final stage of the emotion processing framework translates the computed emotional state into observable robotic behavior. The robot's facial expression and associated behavioral cues are generated using an emotion expression mapping function defined as

$$F_{expr} = f_{face}(\alpha, \lambda) \tag{10}$$

This mapping function encodes both the categorical emotion type α and its corresponding intensity λ, enabling the robot to produce graded and contextually appropriate affective expressions.

By incorporating both emotion category and intensity, the system allows variations in expressive behavior even under similar conversational contexts. The final behavioral output is therefore influenced by the interaction of internal emotion dynamics, external sensory stimuli, and personality-modulated adaptation parameters.

V. Hardware Prototype

The hardware prototype as shown in Fig. 3 was developed to physically embody the proposed emotional interaction framework and to provide an integrated platform for multimodal sensing and expressive communication. The system architecture consists of a microcontroller connected to a laptop that serves as an edge computing unit for executing the Large Language Model (LLM). The laptop also provides the primary audio interface, utilizing its built-in microphone for capturing user speech and its speaker for delivering verbal responses.

To capture user interaction, the prototype integrates three sensing modalities. A tactile sensor enables direct touch detection, a range sensor monitors user proximity and movement, and an inertial measurement unit (IMU) tracks orientation and motion dynamics. These multimodal sensing components support the extraction of affective and contextual interaction cues.

For emotional expression, the robot incorporates two output mechanisms. A display screen presents visual emotional cues corresponding to computed affective states, while a haptic motor provides vibratory feedback that reinforces emotional responses through physical interaction.

The physical design adopts a compact, toy-inspired form factor to reduce the mechanical and industrial appearance typically associated with robotic systems. This design approach aims to enhance user comfort, accessibility, and natural engagement during interaction.

Through this integrated configuration, the hardware platform enables the robot to process multimodal user inputs and to generate synchronized verbal and non-verbal emotional responses. This implementation bridges the gap between computational emotion modeling and embodied human–robot interaction, supporting real-time emotionally adaptive communication.

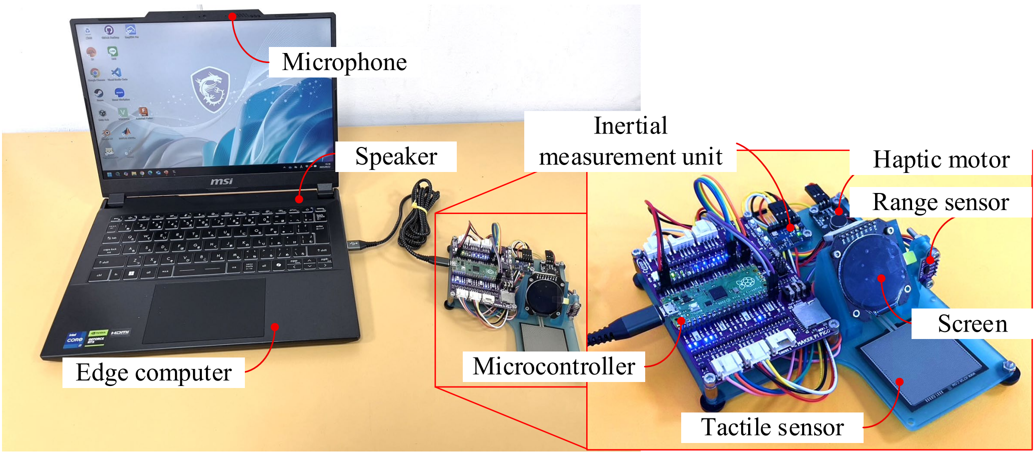


Fig. 3. Robotic interface prototype.

VI. Experimental Setup

This section describes the experimental design used to validate the proposed emotional interaction framework. The experiments aim to verify the feasibility of emotion extraction from sensor inputs and to evaluate how personality parameters influence affective adaptation in human–robot interaction.

The initial evaluation focuses primarily on touch-based interaction, as physical contact provides a direct and intuitive channel for conveying user intent and affective cues. The present evaluation considers both voice and tactile modalities, while additional sensing components integrated in the robotic prototype are reserved for future extensions. In the experiment, predefined touch patterns are mapped to specific emotional categories. For example, gentle tapping is associated with positive emotional activation, whereas abrupt or stronger contact is mapped to negative emotional responses. These mappings are predefined and stored in a reference table that links sensor readings with emotional type and intensity.

To examine personality-driven adaptation, the system is tested across different combinations of personality parameters, namely sensitivity and consideration levels. Sensitivity regulates how strongly external stimuli influence the robot's internal emotional state, whereas consideration controls how closely the robot aligns its outward expression with the user's emotional state.

Three performance metrics are used to evaluate the system:

a) Inner emotion values, representing the robot's intrinsic affective state
b) Social emotion values, representing the robot's outward emotional expression
c) Processing time, which reflects system responsiveness during conversational interaction

Together, these metrics enable evaluation of both emotional modeling performance and real-time operational capability.

A. Interaction Processing Flow

Fig. 4 illustrates the overall interaction processing pipeline of the proposed system. The process begins when the user provides input through speech or physical interaction. Speech input is analyzed using the text analytics module, which extracts semantic content and topic emotion from the utterance. Simultaneously, external stimuli such as tactile interaction contribute stimulus–response emotion derived from sensor input.

These emotional components are integrated to compute the robot's inner emotion, representing its intrinsic affective state. The inner emotion is then used to update personality-modulated parameters and to support subsequent emotional processing. The system then calculates social emotion, which regulates outward emotional expression by balancing internal emotional state with consideration of the user's emotional condition.

The final response is generated by integrating contextual information, semantic content, and emotional alignment. The output is expressed through both verbal responses and behavioral cues. As demonstrated in Fig. 4, a user expressing positive travel experiences leads the robot to generate supportive and emotionally aligned responses, illustrating the framework's ability to support natural and socially coherent interaction.

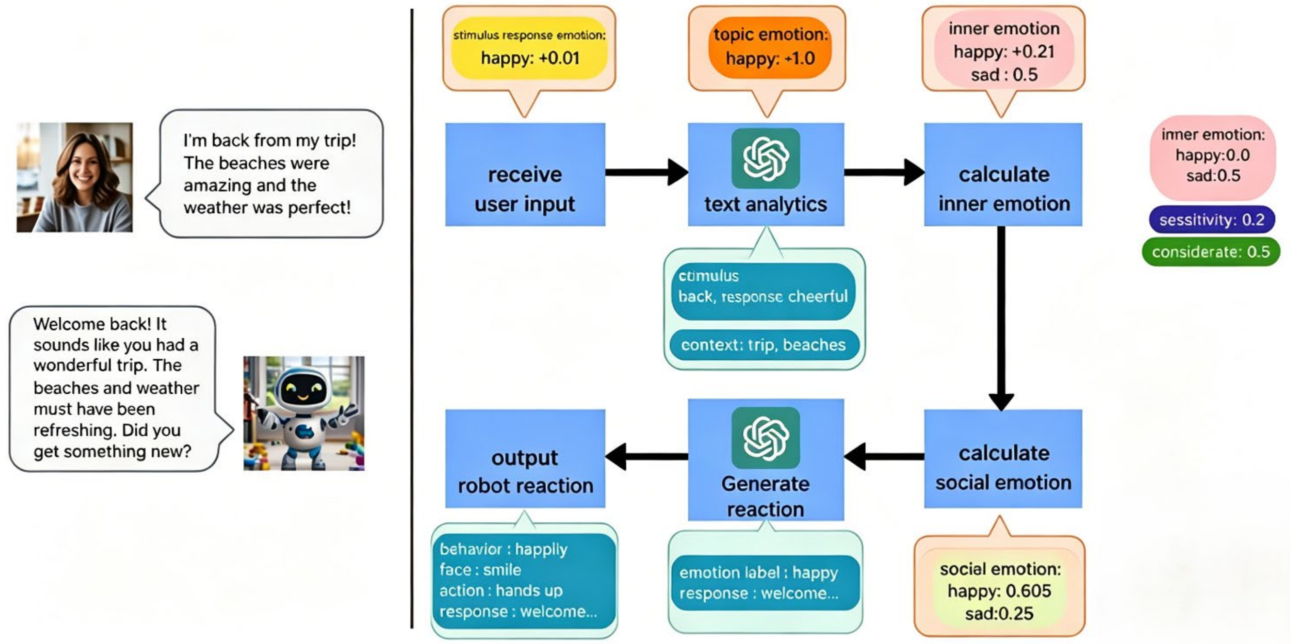


Fig. 4. Example interaction flow diagram.

B. Influence of Personality Parameters

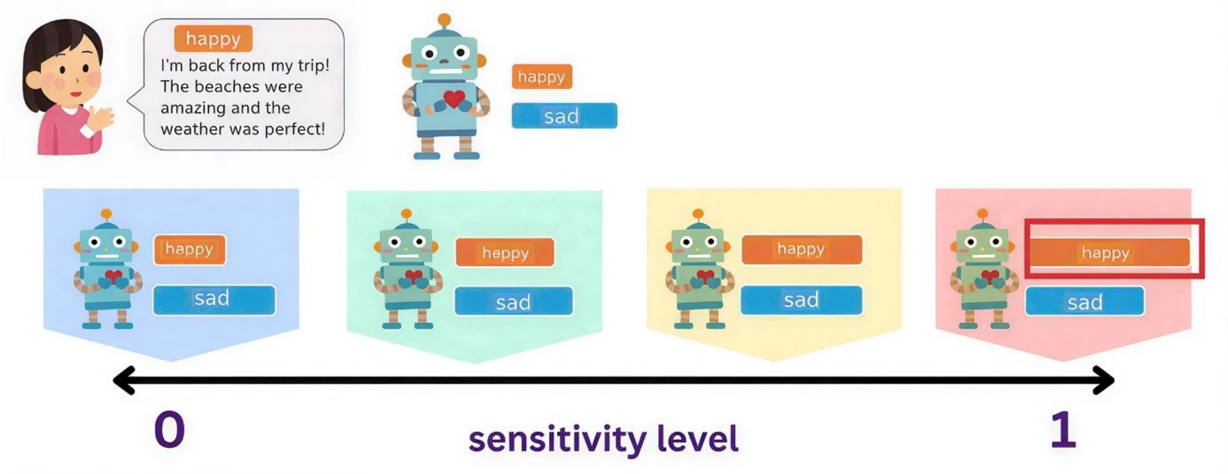
Fig. 5 demonstrates the influence of personality parameters on emotional adaptation and behavioral response. Sensitivity is defined as the degree to which the robot's emotional state is influenced by external emotional stimuli.

In the illustrated example, the robot initially exhibits a negative emotional state while the user expresses positive emotion. When sensitivity is set to zero, the robot's internal emotional state remains unchanged, indicating complete emotional stability. At intermediate sensitivity levels, the robot partially adapts to the user's emotion while maintaining aspects of its prior emotional state. At maximum sensitivity, the robot fully aligns its internal emotional state with the user's affective expression.

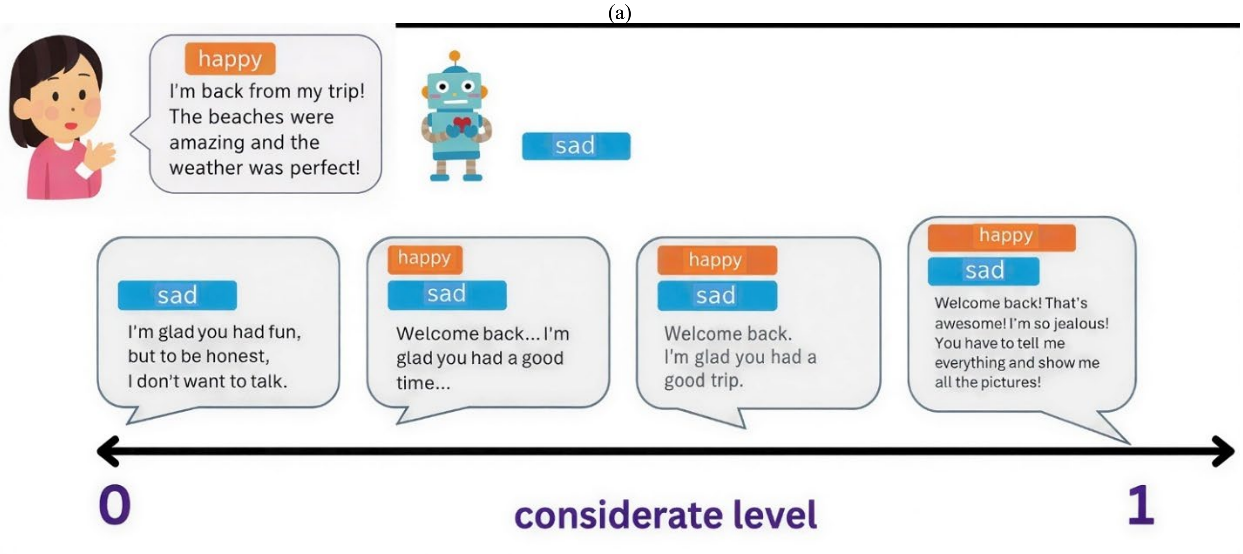
These results illustrate that sensitivity acts as a key parameter governing the trade-off between emotional stability and emotional adaptability, demonstrating the importance of personality modulation in affective human–robot interaction.

C. Emotional State Transition and Decision Mechanism

Fig. 6 presents the method used to determine the robot's current emotional state. In the proposed framework, emotions are represented as numerical intensity values associated with categorical labels. The dominant emotional state is selected as the label with the highest intensity value.



(a)

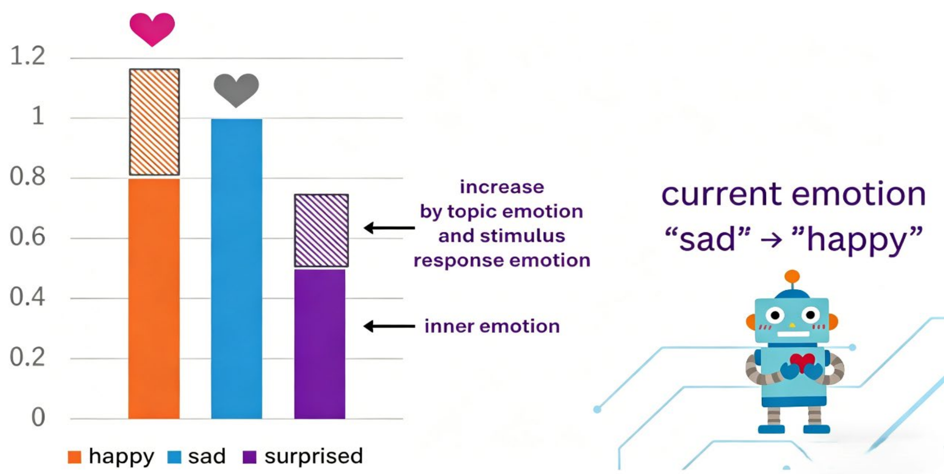


(b)

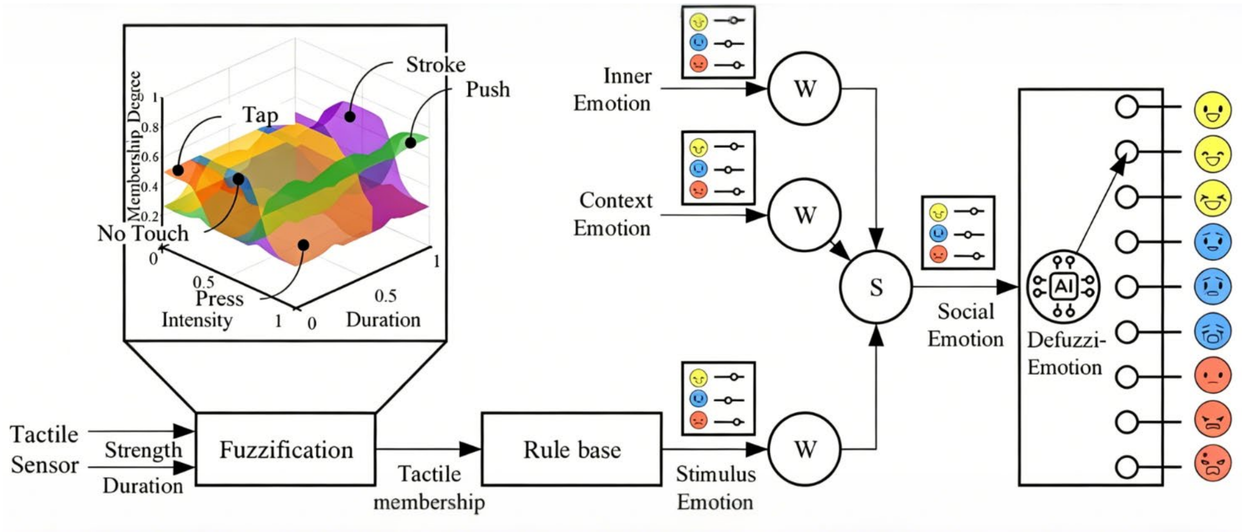
Fig. 5. Conceptual illustration of the influence of sensitivity and consideration parameters on emotional response: (a) Effect of personality to emotion and behavior, (b) Effect of personality to emotion and behavior.

In the example shown in Fig. 6, the robot initially exhibits emotional values of 0.8 for happy, 1.0 for sad, and 0.5 for surprised. Since sadness has the highest value, the robot is initially classified as sad. After integrating new emotional inputs, the value of happiness exceeds that of sadness, resulting in an emotional state transition.

This mechanism enables continuous emotional updating and supports adaptive behavioral responses based on both interaction history and external stimuli.



(a)



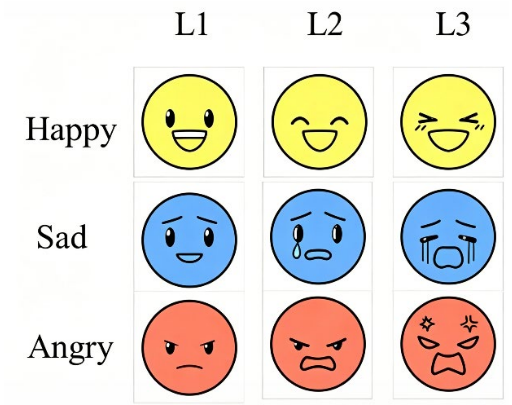
(b)

Fig. 6. Emotional calculations and decisions: (a) The resulting social emotion, (b) Emotional flow diagram.

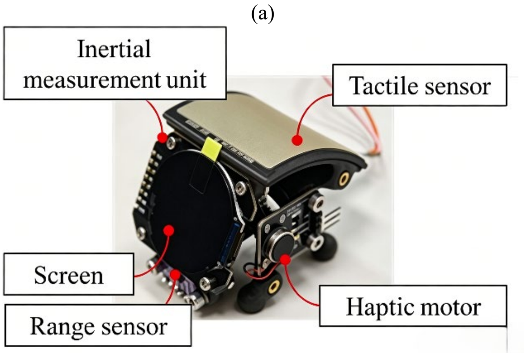
D. Visual Emotional Expression

Fig. 7 illustrates the visual emotional expressions implemented in the robotic prototype. The robot uses a display screen to present facial expressions corresponding to emotional categories including happy, sad, and angry. Each category includes multiple expression levels representing different emotional intensities.

The use of graded facial expressions improves emotional transparency and allows users to interpret the robot's affective state intuitively. This design enables subtle emotional variations to be communicated effectively, thereby enhancing naturalness and social acceptability during interaction.



(a)



(b)

(c)

Fig. 7. Robot faces: (a) Facial emotion. (b) The robot's core hardware. (c) The robot exterior.

Fig. 7(c) presents an example evaluation in which a stroking interaction is predefined as a positive emotional stimulus with intensity +1.0. In this experiment, sensitivity and consideration parameters are both set to 0.5 to achieve balanced emotional adaptation.

The results indicate that the inner emotion value associated with happiness increases to 0.25, while the corresponding social emotion reaches 0.625. Emotional values for sadness and anger remain negligible, resulting in a positive emotional classification. Based on these values, the robot generates a smiling facial expression displayed on the interface.

This experimental result demonstrates that the system can successfully extract emotional information from sensor input, modulate emotional states through personality parameters, and generate emotionally consistent expressions in real time.

VII. Results

The experimental evaluation demonstrates that the proposed emotional interaction framework is capable of generating emotionally adaptive responses from multimodal user input. The evaluation includes three complementary analyses: an emotion transition experiment examining the temporal evolution of emotional states, a response latency experiment measuring system responsiveness, and a parameter study analyzing the influence of personality-modulated sensitivity and consideration values.

In the conducted test cases, the system successfully extracted emotional information from both voice and tactile signals, computed corresponding inner and social emotional states, and generated contextually appropriate responses using the Large Language Model (LLM). The results confirm that personality parameters significantly influence emotional adaptation. As conceptually illustrated in Fig. 5(a) and 5(b), variations in sensitivity influence how strongly user emotional input affects the robot's internal emotional state. Lower sensitivity values preserve emotional stability by maintaining the influence of previous emotional states, whereas higher sensitivity values allow the system to respond more rapidly to external affective stimuli. Similarly, the consideration parameter regulates the balance between internally driven emotional tendencies and outward expressions aligned with the user's emotional context. The illustrative numerical examples presented in Tables IV and V further demonstrate how different parameter settings influence the resulting emotional state and expression behavior under controlled interaction conditions.

The experimental results further demonstrate that identical contextual and semantic inputs can produce different emotional responses depending on the interaction history, emotional state values, and personality parameter settings. This variability indicates that the proposed framework supports adaptive and dynamic emotional response generation during interaction.

In addition, response latency measurements show that the system can perform emotional computation and response generation within a time range suitable for near real-time conversational interaction. The measured latency values under both voice and tactile interactions indicate that the system maintains stable responsiveness during experimental trials.

A. Emotion Transition Experiment

To illustrate the temporal behavior of the proposed emotional framework, a controlled sequential interaction simulation was conducted. The objective of this experiment was to analyze how the internal emotional state evolves in response to different interaction stimuli. The simulation was organized as a sequence of discrete interaction steps, where each step represents a single interaction event between the user and the robotic system. These steps are denoted as $t_0, t_1, t_2, \ldots$ where each step corresponds to the moment when a new interaction stimulus is applied and the emotional state of the system is updated. At each interaction step, only one input modality was activated in order to avoid overlapping stimuli and to clearly observe the influence of each interaction type on the emotional state.

Two interaction modalities consistent with the proposed system architecture were considered as inputs: voice and touch. Voice inputs were represented by short textual utterances corresponding to speech recognition outputs. These utterances included positive phrases (e.g., "Good job", "Thank you"), neutral expressions (e.g., "Hello"), and negative phrases (e.g., "Leave me alone", "I am disappointed"). Touch interactions were simulated using predefined tactile patterns including tap, stroke, press, and push. These interaction types were categorized according to their emotional polarity, where tap and stroke represent positive interactions, while press and push correspond to negative interactions. Each interaction stimulus was mapped to an emotional stimulus vector influencing the robot's internal emotional state.

The simulation was initialized with a neutral emotional state to avoid bias toward any specific emotion. The initial emotion vector was defined with approximately balanced values across the three emotional dimensions: happy, sad, and angry. Starting from this neutral state allows the influence of subsequent interaction stimuli to be observed more clearly as the emotional state evolves during the interaction sequence.

At each interaction step, the emotional state of the system was updated using a temporal emotion update rule. The updated emotion vector was computed as a weighted combination of the previous emotional state and the stimulus-induced emotional response. This mechanism preserves emotional continuity while allowing external interactions to influence the system. In addition, when no new stimulus was applied, a mild decay mechanism gradually moved the emotional state toward a neutral equilibrium. This decay behavior prevents emotional saturation and reflects the dynamic characteristics of affective interactions over time.

After updating the emotional state, the dominant emotion was determined by selecting the emotional component with the highest value among the three emotional dimensions. The resulting emotion vector represents the current affective state of the system, and the dominant emotion was used to determine the corresponding facial expression displayed by the robot.

The computed emotion scores were subsequently mapped to visual emotional expressions on the robot interface. Each emotional category (happy, sad, and angry) was associated with three levels of expression intensity. The expression level was determined based on the magnitude of the dominant emotional score, where higher scores correspond to stronger facial expressions. This mapping enables the system to represent graded emotional responses rather than simple binary emotional states.

By sequentially applying the interaction stimuli and updating the emotional state at each interaction step, a temporal sequence of emotional scores was obtained. The resulting data are summarized in Table II, which presents the interaction step, the applied input stimulus (voice or touch), and the corresponding emotional scores for happy, sad, and angry. The table illustrates how the emotional state dynamically evolves in response to positive, negative, and neutral interactions, thereby demonstrating the behavior of the proposed emotional framework under controlled interaction scenarios.

Each interaction step ($t_0, t_1, t_2, \ldots$) represents a single interaction stimulus applied to the system, and the resulting emotional scores (happy, sad, angry) represent the updated internal emotional state of the robot.

Fig. 8 illustrates the temporal variation of the three emotional dimensions (happy, sad, and angry) during the interaction sequence described in Table II. Each step corresponds to a single interaction event involving either voice or touch stimuli.

The figure shows that positive interaction stimuli gradually increase the happy emotion score, while negative stimuli lead to transitions toward sad or angry emotional states. The curves also demonstrate the effect of emotional continuity in the proposed framework, where previous emotional states influence subsequent emotional responses. As a result, emotional transitions occur progressively across interaction steps rather than instantaneously, reflecting the temporal dynamics modeled in the proposed emotional state update mechanism.

TABLE II. EMOTION TRANSITION SIMULATION UNDER SEQUENTIAL VOICE AND TOUCH INTERACTION EVENTS

Time	Modality	Voice input	Touch input	Polarity	Happy	Sad	Angry	Dominant emotion	Face level
t0	None	–	–	Neutral	0.33	0.34	0.33	Sad	Sad-L1
t1	Voice	"Hello"	–	Neutral-positive	0.40	0.32	0.28	Happy	Happy-L1
t2	Voice	"Good job"	–	Positive	0.52	0.27	0.21	Happy	Happy-L2
t3	Touch	–	Tap	Slightly positive	0.58	0.23	0.19	Happy	Happy-L2
t4	Touch	–	Stroke	Positive	0.69	0.18	0.13	Happy	Happy-L2
t5	Voice	"Thank you"	–	Positive	0.75	0.14	0.11	Happy	Happy-L3
t6	None	–	–	Neutral decay	0.68	0.18	0.14	Happy	Happy-L2
t7	Voice	"I feel tired"	–	Mild negative	0.56	0.29	0.15	Happy	Happy-L2
t8	Voice	"I am disappointed"	–	Negative	0.42	0.42	0.16	Happy/Sad tie	Transition
t9	Touch	–	Press	Negative	0.31	0.47	0.22	Sad	Sad-L1
t10	Touch	–	Push	Strong negative	0.20	0.44	0.36	Sad	Sad-L1
t11	Voice	"Leave me alone"	–	Negative-angry	0.14	0.36	0.50	Angry	Angry-L2
t12	None	–	–	Neutral decay	0.18	0.34	0.48	Angry	Angry-L1
t13	Voice	"Calm down"	–	Soothing positive	0.30	0.30	0.40	Angry	Angry-L1
t14	Touch	–	Stroke	Positive	0.46	0.24	0.30	Happy	Happy-L1
t15	Voice	"It's okay"	–	Positive-soothing	0.58	0.18	0.24	Happy	Happy-L2
t16	Voice	"You did well"	–	Positive	0.66	0.15	0.19	Happy	Happy-L2
t17	Touch	–	Tap	Slightly positive	0.71	0.12	0.17	Happy	Happy-L3
t18	None	–	–	Neutral decay	0.64	0.16	0.20	Happy	Happy-L2
t19	Voice	"Why did you do that?"	–	Negative	0.50	0.24	0.26	Happy	Happy-L2
t20	Touch	–	Press	Negative	0.38	0.35	0.27	Happy	Happy-L1
t21	Voice	"Stop it"	–	Strong negative	0.25	0.29	0.46	Angry	Angry-L1
t22	Touch	–	Push	Strong negative	0.16	0.24	0.60	Angry	Angry-L2
t23	Voice	"Let's be friends"	–	Positive	0.35	0.22	0.43	Angry	Angry-L1
t24	Touch	–	Stroke	Positive	0.54	0.16	0.30	Happy	Happy-L2

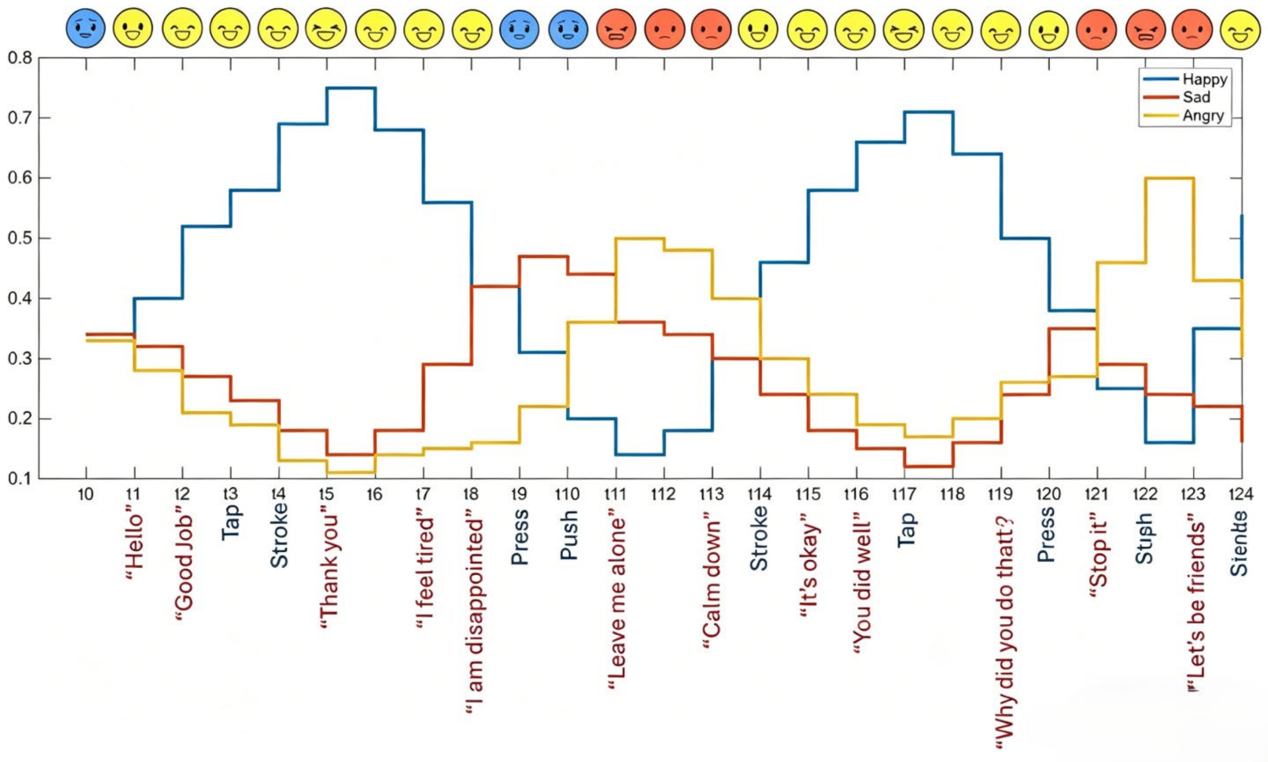


Fig. 8. Emotion score evolution across interaction steps.

B. Response Latency Experiment

To evaluate the responsiveness of the proposed emotional interaction framework, a response latency experiment was conducted under two primary interaction modalities: voice and touch. The objective of this experiment was to measure the end-to-end response time of the system, defined as the time elapsed from the moment an interaction event is detected to the moment the corresponding emotional expression is rendered on the robot interface.

The system follows an edge–server architecture. The experiment was conducted using a local server connected to the robotic prototype through a wireless network. A microcontroller-based edge controller detects interaction events and forwards the corresponding input data to the server. The server executes the emotion processing module, which interprets the input stimulus, updates the internal emotional state, determines the dominant emotion, and generates the corresponding facial expression output. The computed emotional result is then transmitted back to the edge controller, which renders the final expression on the robot interface. Consequently, the measured latency includes event detection at the edge controller, communication to the server, server-side emotion computation, return transmission, and final expression rendering.

Two categories of interaction events were evaluated in this experiment. For the voice modality, representative utterances with different emotional polarity were used, including “Hello,” “Good job,” and “Leave me alone.” For the touch modality, predefined tactile interactions including tap, stroke, press, and push were used as representative stimuli.

Each interaction type was evaluated through 10 repeated trials in order to account for variability in system timing. For every trial, the response latency was measured as the elapsed time between the detection of the interaction event at the edge controller and the display of the resulting emotional expression after server-side processing. For voice interaction, latency was measured from the moment the recognized voice event was registered at the edge controller, after speech recognition processing. For tactile interaction, latency measurement started when the touch sensor signal was detected by the edge controller. The latency values obtained from repeated trials were then used to compute the mean latency, standard deviation, minimum, and maximum values for each interaction type.

The resulting measurements are summarized in Table III, which provides a quantitative evaluation of the system’s end-to-end response latency under different interaction modalities.

C. Sensitivity and Consideration Parameter Study

To further clarify the role of the personality-modulated parameters in the proposed emotional framework, an illustrative numerical example is presented focusing on the sensitivity and consideration parameters. The objective of this analysis is to demonstrate how these parameters influence the evolution of the internal emotional state and the resulting social emotional expression under controlled interaction conditions. The example is derived from the proposed emotion update mechanism and is intended to illustrate the behavior of the model rather than to provide a full empirical validation.

In the proposed framework, the sensitivity parameter controls the degree to which external stimuli influence the internal emotional state of the system. A higher sensitivity value allows the emotional state to adapt more rapidly to new interaction inputs, while a lower sensitivity value preserves emotional inertia and maintains the influence of the previous emotional state. In the extreme case where the sensitivity value approaches zero, the internal emotional state is dominated by the previous emotional tendency and external stimuli have minimal influence.

Table IV presents an illustrative numerical example showing the effect of sensitivity on emotional state adaptation. In this example, the system is initialized with a relatively negative internal emotional tendency, and a positive voice stimulus (“Good job”) is applied. When the sensitivity value is low, the emotional state remains dominated by the previous negative tendency. As the sensitivity value increases, the influence of the positive stimulus becomes stronger, resulting in a transition toward a happy dominant emotion and a higher facial expression intensity.

The parameter values were selected as representative low, medium, and high operating conditions to illustrate the qualitative behavior of the proposed emotional model.

TABLE III. END-TO-END RESPONSE LATENCY UNDER VOICE AND TOUCH INTERACTION MODALITIES

Interaction type	Input example	Mean latency (ms)	Std. dev. (ms)	Min (ms)	Max (ms)
Voice	“Hello”	812	64	732	914
Voice	“Good job”	845	71	751	962
Voice	“Leave me alone”	892	78	803	1015
Touch	Tap	214	18	188	249
Touch	Stroke	236	21	205	274
Touch	Press	228	19	201	262
Touch	Push	247	23	214	288

TABLE IV. ILLUSTRATIVE NUMERICAL EXAMPLE SHOWING THE EFFECT OF SENSITIVITY ON EMOTIONAL STATE ADAPTATION

Sensitivity	Consideration	Input Stimulus	Happy	Sad	Angry	Dominant Emotion	Face level
0.0	0.5	Voice: “Good job”	0.24	0.58	0.18	Sad	Sad-L2
0.5	0.5	Voice: “Good job”	0.46	0.39	0.15	Happy	Happy-L1
1.0	0.5	Voice: “Good job”	0.71	0.18	0.11	Happy	Happy-L3

In addition to sensitivity, the consideration parameter regulates how the robot balances its internal emotional state with the perceived emotional context of the interaction when generating social emotional expressions. A lower consideration value results in emotional expressions that align more strongly with the user's stimulus, whereas a higher consideration value preserves the influence of the robot's internal emotional tendency.

Table V illustrates the effect of the consideration parameter under a controlled interaction scenario. In this case, a positive voice stimulus ("Thank you") is applied while the internal emotional tendency still contains a negative component. When the consideration value is low, the resulting emotional expression aligns closely with the positive user input. As the consideration value increases, the robot's outward expression becomes increasingly influenced by its internal emotional state, which may lead to the persistence of negative emotional expressions despite positive external stimuli.

Similarly, the consideration parameter values were selected as representative low, medium, and high operating conditions to illustrate their influence on social emotional expression.

TABLE V. ILLUSTRATIVE NUMERICAL EXAMPLE SHOWING THE EFFECT OF CONSIDERATION ON SOCIAL EMOTIONAL EXPRESSION

Sensitivity	Consideration	Input stimulus	Happy	Sad	Angry	Dominant emotion	Face level
0.5	0.0	Voice: "Thank you"	0.69	0.14	0.17	Happy	Happy-L2
0.5	0.5	Voice: "Thank you"	0.47	0.16	0.37	Happy	Happy-L1
0.5	1.0	Voice: "Thank you"	0.26	0.15	0.59	Angry	Angry-L2

These illustrative numerical examples highlight how the personality-modulated parameters influence emotional adaptation and outward expression in the proposed framework. In particular, sensitivity primarily affects the responsiveness of the internal emotional state to new stimuli, while consideration regulates the balance between internal emotion and socially expressed emotion during interaction.

It should be emphasized that these values are intended to illustrate the qualitative behavior of the proposed emotion update mechanism. They serve as explanatory numerical cases demonstrating how parameter variations influence emotional transitions within the model, rather than representing empirical measurements obtained from user-based experiments.

VIII. DISCUSSION

The experimental findings suggest that integrating multimodal affect perception with personality-modulated emotional computation can enhance interaction quality in conversational robotic systems. Compared with conventional sentiment-based conversational models, the proposed dual-layer emotional framework introduces temporal emotional continuity through internal emotion modeling while enabling socially adaptive outward expression through social emotion regulation.

To further clarify the advantages of the proposed framework, Table VI compares the dual-layer emotional architecture with conventional reactive emotion systems and single-layer affective models. Unlike reactive emotion systems that directly map user input to emotional responses, the proposed framework introduces an internal emotional layer that maintains emotional continuity across interactions. In addition, the separation between internal emotion and social emotion enables personality-modulated adaptation and socially coherent behavior. These characteristics provide improved flexibility and emotional consistency compared with conventional affective interaction systems.

TABLE VI. COMPARISON WITH CONVENTIONAL EMOTIONAL INTERACTION FRAMEWORKS

Feature	Reactive emotion system	Single-layer emotion model	Proposed dual-layer framework
Internal emotion memory	No	Limited	Yes
Social emotion adaptation	No	Limited	Yes
Personality modulation	No	Rare	Yes
Multimodal interaction	Limited	Limited	Yes
Emotional continuity	No	Partial	Yes
Implementation on robot	Limited	Limited	Yes

The ability to generate different emotional responses from identical conversational content highlights the system's capacity to simulate human-like affective variability. This capability is particularly relevant in human–robot interaction, where emotional congruence contributes to user engagement, trust, and perceived naturalness.

The results also indicate that personality parameters serve as an effective mechanism for regulating emotional adaptability. By adjusting sensitivity and consideration levels, the system can generate diverse interaction styles ranging from emotionally stable to more responsive affective behaviors. This flexibility supports potential personalization and long-term interaction scenarios.

However, the current evaluation primarily focuses on controlled interaction simulations and system-level latency measurements. While these results illustrate the behavior of the proposed emotional framework, further studies involving user-based evaluations and larger-scale interaction experiments will be necessary to fully assess the perceived effectiveness of the emotional responses in real human–robot interaction settings.

In addition, real-time performance remains an important design consideration. Although the current system achieved responsive interaction during the experimental tests, further optimization of emotional computation and LLM processing latency will be beneficial for deployment in continuous real-world interaction environments.

IX. Conclusions

This study presented a dual-layer emotional framework for human–robot conversational interaction that integrates internal emotion dynamics and socially expressive emotion within a personality-modulated architecture. The proposed framework illustrates how personality parameters, particularly sensitivity and consideration, can regulate the balance between emotional stability and social adaptability, enabling robotic agents to generate emotionally coherent and socially appropriate responses.

Through the development of a hardware prototype and system-level evaluation, the study demonstrates the feasibility of combining multimodal sensing, layered emotional computation, and large language model (LLM) integration within an embodied conversational robotic system. The experimental results illustrate that the framework can support adaptive emotional behavior beyond conventional reactive sentiment-based conversational approaches.

The findings indicate that personality-driven emotional modulation provides a flexible mechanism for adjusting interaction behavior. By maintaining internal emotional continuity while adapting outward emotional expression, the proposed framework enables conversational agents to produce context-dependent emotional responses during interaction. From a Human–Computer Interaction (HCI) perspective, the framework offers a system-level approach for designing emotionally adaptive multimodal conversational interfaces.

Future work will focus on extending the framework to support multi-user and long-term interaction scenarios in order to evaluate emotional consistency and personalization during extended interactions. Additional sensing modalities, including facial expression analysis and physiological signal monitoring, will be explored to improve robustness in emotion perception. Further user-based evaluations will also be conducted to assess the perceived effectiveness of the emotional interaction model in real human–robot interaction settings.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Shitara Kaede, Kosuke Takano, and Ratchatin Chancharoen conceptualized the study; Shitara Kaede, Kantawatchr Chaiprabha, Pimolkan Piankitrungrean, and Kosuke Takano developed the methodology; Shitara Kaede, Kantawatchr Chaiprabha, Pimolkan Piankitrungrean, Kosuke Takano, and Ratchatin Chancharoen performed validation and formal analysis; Shitara Kaede and Kantawatchr Chaiprabha conducted the investigation; Shitara Kaede curated the data; Shitara Kaede, Kantawatchr Chaiprabha, and Pimolkan Piankitrungrean prepared visualizations; Shitara Kaede wrote the original draft; Shitara Kaede, Kosuke Takano, Ratchatin Chancharoen, and Gridsada Phanomchoeng reviewed and edited the manuscript; Kosuke Takano and Ratchatin Chancharoen supervised the project; Kosuke Takano, Ratchatin Chancharoen, and Gridsada Phanomchoeng acquired funding and provided resources; all authors approved the final version of the manuscript.

Funding

This research project is supported by the Second Century Fund (C2F), Chulalongkorn University, Bangkok, Thailand and 111th Anniversary Engineering Research Catalyst Fund Towards U Top 100.

Acknowledgment

The authors gratefully acknowledge the support provided by the IMT Lab, Faculty of Engineering, Chulalongkorn University.

References

[1] Y. Zhang, S. Sun, M. Galley, Y. C. Chen, C. Brockett, X. Gao, J. Gao, J. Liu, and W. B. Dolan, "DialoGPT: Large-scale generative pre-training for conversational response generation," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics: Syst. Demonstrations*, July 2020, pp. 270–278.

[2] S. U. Singh and A. S. Namin, "A survey on chatbots and large language models: Testing and evaluation techniques," *Natural Language Processing Journal*, 100128, 2025.

[3] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Applications*, vol. 2, 100006, 2020.

[4] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science Robotics*, vol. 3, no. 21, 5954, 2018.

[5] T. B. Sheridan, "A review of recent research in social robotics," *Current Opinion in Psychology*, vol. 36, pp. 7–12, 2020.

[6] M. Shiomi, H. Sumioka, and H. Ishiguro, "Survey of social touch interaction between humans and robots," *J. Robotics and Mechatronics*, vol. 32, no. 1, pp. 128–135, 2020.

[7] C. Esterwood, K. Essenmacher, H. Yang, F. Zeng, and L. P. Robert, "A meta-analysis of human personality and robot acceptance in human–robot interaction," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2021, pp. 1–18.

[8] C. Esterwood and L. P. Robert, "A systematic review of human and robot personality in health care human–robot interaction," *Frontiers in Robotics and AI*, vol. 8, 748246, 2021.

[9] Y. Wu, Q. Mi, and T. Gao, "A comprehensive review of multimodal emotion recognition: Techniques, challenges, and future directions," *Biomimetics*, vol. 10, no. 7, 418, 2025.

[10] B. Pan, K. Hirota, Z. Jia, and Y. Dai, "A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods," *Neurocomputing*, vol. 561, 126866, 2023.

[11] C. Clabaugh *et al.*, "Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders," *Frontiers in Robotics and AI*, vol. 6, 110, 2019.

[12] L. P. Robert *et al.*, "A review of personality in human–robot interactions," *Foundations and Trends in Information Systems*, vol. 4, no. 2, pp. 107–212, 2020.

[13] H. Lian, C. Lu, S. Li, Y. Zhao, C. Tang, and Y. Zong, "A survey of deep learning-based multimodal emotion recognition: Speech, text, and face," *Entropy*, vol. 25, no. 10, 1440, 2023.

[14] A. A. Gladys and V. Vetriselvi, "Survey on multimodal approaches to emotion recognition," *Neurocomputing*, vol. 556, 126693, 2023.

[15] Y. H. H. Tsai *et al.*, "Multimodal transformer for unaligned multimodal language sequences," in *Proc. Annu. Meeting Assoc. Comput. Linguistics*, July 2019.
[16] R. Das and T. D. Singh, "Multimodal sentiment analysis: A survey of methods, trends, and challenges," *ACM Comput. Surveys*, vol. 55, no. 13, pp. 1–38, 2023.
[17] T. Baltruaitis, C. Ahuja, and L. P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019.
[18] N. J. Shoumy *et al.*, "Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals," *J. Network Comput. Appl.*, vol. 149, 102447, 2020.
[19] S. Roller *et al.*, "Recipes for building an open-domain chatbot," in *Proc. 16th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2021, pp. 300–325.
[20] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27730–27744, 2022.
[21] M. P. A. Ramaswamy and S. Palaniswamy, "Multimodal emotion recognition: A comprehensive review, trends, and challenges," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 14, no. 6, e1563, 2024.
[22] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: A survey," *Int. J. Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.
[23] M. Cansev *et al.*, "Implementing social and affective touch to enhance user experience in human–robot interaction," *Frontiers in Robotics and AI*, vol. 11, 1403679, 2024.
[24] G. Laban *et al.*, "Building long-term human–robot relationships: Examining disclosure, perception and well-being across time," *Int. J. Social Robotics*, vol. 16, no. 5, pp. 1–27, 2024.
[25] M. Staffa, L. D'Errico, and A. Maratea, "Influence of social identity and personality traits in human–robot interactions," *Robotics*, vol. 13, no. 10, 144, 2024.
[26] K. Kabacińska *et al.*, "Influence of user personality traits and attitudes on interactions with social robots: A systematic review," *Collabra: Psychology*, vol. 11, no. 1, 129175, 2025.
[27] R. Stock-Homburg, "Survey of emotions in human–robot interactions: Perspectives from robotic psychology on 20 years of research," *Int. J. Social Robotics*, vol. 14, no. 2, pp. 389–411, 2022.
[28] M. Okuda, Y. Takahashi, and S. Tsuichihara, "Human response to humanoid robot that responds to social touch," *Applied Sciences*, vol. 12, no. 18, 9193, 2022.
[29] K. Matheus, R. Ramnauth, B. Scassellati, and N. Salomons, "Long-term interactions with social robots: Trends, insights, and recommendations," *ACM Trans. Human–Robot Interaction*, vol. 14, no. 3, pp. 1–42, 2025.