Dynamic Modulation of Emotional Expressions in Social Robots: Effects on Liveliness and Naturalness

Haeun Park[1], Sun Jun Hwang[2], Hyojin Kim[3], Jiyeon Lee[2], Hui Sung Lee[2]

***Abstract*— Humans naturally express emotions with subtle variations, and exaggerated expressions often appear as heightened intensity in facial, bodily, or vocal cues. This paper introduces a method for exaggerating robotic emotional expressions by dynamically adjusting intensity within an emotion dynamics model. By systematically manipulating the damping ratio, we generated five distinct intensity levels for each emotion, thereby producing emotional expressions that exhibited different degrees of overshoot. A user study revealed that liveliness ratings for surprise increased linearly with intensity, suggesting that exaggerated, high-energy dynamics are particularly effective for conveying surprise. In contrast, other emotions exhibited optimal points at intermediate levels, indicating that excessive exaggeration can reduce perceived naturalness. These findings highlight the need for emotion-specific and user-specific calibration of expression intensity, supporting more nuanced and engaging human-robot interactions.**

I. INTRODUCTION

Social robots are increasingly being designed to engage in long-term interactions with humans by conveying emotions through multiple modalities, such as facial expressions, body movements, and nonverbal sounds. Emotional expression plays a central role in building trust, fostering empathy, and sustaining user engagement over time [2], [3]. A wide range of computational emotion models have been studied to support such expressions [4]–[7], but most efforts have either focused solely on the generation of emotions or on making robots appear as if they possess emotions. However, in human communication, emotions are often expressed dynamically—for instance, joyful states are often accompanied by more energetic gestures, and surprise reactions may be exaggerated with overshooting movements. This highlights the need for mechanisms that can regulate the dynamics of emotional expression. Nevertheless, relatively little attention has been given to the role of dynamic exaggeration in modulating the perceived intensity of specific emotions.

Fig. 1. Emotional expression of the prototype robot. The yellow arrows visualize the movement patterns for six emotions, with thickness indicating speed.

To address this gap, we propose a method for exaggerating emotional expression by modulating dynamic intensity within the same emotion. Specifically, we implement an emotion engine on a second-order dynamics model, which enables systematic manipulation of the damping ratio and thus variation in the degree of overshoot. We apply this approach to a prototype robot (see Fig. 1) capable of expressing emotions through facial displays, motion, and non-verbal sounds. A user study was conducted in which participants evaluated Ekman's six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) [8] expressed at five dynamic levels. Participants rated liveliness and naturalness, complemented by post-task interviews that revealed preferences and perceived mismatches across modalities. This design allowed us to investigate whether stronger dynamics consistently enhance perception or whether emotion-dependent and user-dependent "sweet spots" emerge.

II. RELATED WORK

A. Computational emotion models

Computational emotion models generate appropriate emotions in robots based on stimuli [4]. Most approaches assume that a robot possesses an internal emotion or affect space, often derived from psychological theories or defined as the

*This study was partly supported by the Technology Innovation Program (20015056, Commercialization design and development of Intelligent Product-Service System for personalized full silver life cycle care) funded by the Ministry of Trade, Industry & Energy(MOTIE, Korea) and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2020R1F1A1066397). This work is based on the first author's thesis [1]. The authors used Gemini 3.1 Pro for English language editing and grammar enhancement.

[1]Haeun Park is with Creative Design Engineering Graduate School, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea haeunpark@unist.ac.kr

[2]Sun Jun Hwang, Jiyeon Lee and Hui Sung Lee are with Design Department, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea {sunjoon020, delay0320, huisung.lee}@unist.ac.kr

[3]Hyojin Kim is with the Department of Electrical and Electronic Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea. kim24@unist.ac.kr

robot's own. Representative models include the AVS space [9], Russell's PAD model [10], [11], and Plutchik's emotion wheel [12]. Once such spaces are defined, computations map external stimuli onto the axes to determine emotional states. For instance, PAD has been widely applied to associate stimuli with emotional values [13], [14], while other works exploit Plutchik's wheel with quantum circuits [15] or apply dynamic shifts within a robot's own affect space [16], [17].

B. Expressive robots and dynamic expression

Early social robots established that emotion expression increases likability and engagement [2], [18]. Subsequent HRI studies showed that motion timing, speed, and smoothness shape perceived affect, animacy, and intent [19]–[21]. Parameterized motion frameworks further link low-level kinematics to human perception, supporting controllable intensity and oscillation [22]. Our approach only uses the *damping ratio* while preserving rise time, enabling a test of how dynamic modulation influences perceived *liveliness* vs. *naturalness*.

Insights from animation theory provide an additional lens. The notion of overshoot aligns with Disneys principle of Follow Through and Overlapping Action, in which movements extend beyond their target before settling, thereby conveying realism. At the same time, it resonates with the principle of Exaggeration, where motion is intentionally amplified to heighten expressiveness. In robotic emotion expression, controlled overshoot can introduce naturalistic dynamical effects, while calibrated exaggeration enhances the vividness and clarity of conveyed emotions.

C. Multimodal emotional expression on robots

Facial displays remain a primary channel for conveying emotion, often guided by facial action coding system and animation-inspired control points (CPs) [18], [23]. Multimodal integration (face, body, sound) tends to improve believability when channels are consistent, but cross-modal mismatches can reduce perceived naturalness.

Non-verbal audio can efficiently signal arousal and valence through pitch, dynamics, and spectral structure [24], [25]. In robotics, stylized sound effects complement visual/motion cues when speech is unnecessary or undesirable. We map stimulus-driven triggers to sonic envelopes and carrier/modulator relationships, synchronizing with entry into emotion regions in affect space.

III. EMOTION ENGINE

In this section, the dynamics of emotions are described based on Miwa's Mental dynamic [16] and Linear Dynamic Affect-Expression Model [17], [26]. The employment of emotion dynamics has been demonstrated as a method to achieve a more animation-like dynamic in instances when the robot seeks to express its emotions.

A. Stimulus interpretation

The stimulus interpretation module analyzes incoming sensory stimuli and assigns each sensor a cumulative value x_j ($x_j \in \mathbb{R}$ and $-1 \leq x_j \leq 1$), allowing the robot to respond adaptively to repeated stimuli based on prior sensory information. Building upon the emotion dynamics framework established in our previous work [27], the emotion amplitude a_i is derived from x_j as follows:

$$\begin{aligned} a_i(t) = \sum_{j=1}^{H} w_{ij} \exp\left(-\frac{\big(x_j(t)-\mu_{ij}\big)^2}{2\sigma_{ij}^2}\right), i = 1 \cdots N \\ \mathbf{a} \triangleq [a_1\ a_2\ \cdots\ a_i\ \cdots\ a_N]^T \\ \sum_{j=1}^{H} w_{ij} = 1, \quad 0 \leq w_{ij} \leq 1 \end{aligned} \tag{1}$$

where μ_{ij} and σ_{ij} represent the center and spread of the Gaussian distribution for the i-th emotion relative to the j-th stimulus, respectively. In this work, the parameters—$\mu_i = \{-0.3, -0.6, -1.0, 1.0, -0.8, 0.5\}$ and $\sigma_i = \{0.1, 0.1, 0.1, 0.2, 0.1, 0.2\}$—were empirically determined to reflect the emotional response characteristics of the system. The amplitude a_i is then used to generate a time-dependent stimulus response $r_i(t)$:

$$\begin{aligned} r_i(t) = \Big(u(t - t_{\text{last}}) - u(t - t_{\text{last}} - T_d)\Big) a_i \\ t_{\text{last}} = \arg\max_{t_k < t}\ t_k \end{aligned} \tag{2}$$

where $u(t)$ denotes the unit step function, t_{last} represents the timestamp of the most recent stimulus event, and T_d is the predefined duration for which the emotion stimulus remains active. This ensures that the stimulus acts as a temporary impulse.

These individual responses are then assembled to construct the force element vector $\mathbf{r}$, which drives the robot's emotional state in the emotion dynamics module:

$$\begin{aligned} \mathbf{r} \triangleq [r_1\ r_2\ \cdots\ r_i\ \cdots\ r_N]^T \\ r_i \geq 0, \quad \|\mathbf{r}\|_1 \leq 1 \end{aligned} \tag{3}$$

Vector $\mathbf{r}$ is an important factor in determining the target coordinates within the affect space. In (3), N denotes the number of basic emotions used in the emotion engine. The l^1-norm constraint is specifically chosen to constrain the total energy of the emotional expression across all modalities, thereby influencing the robot's affective response.

B. Emotion dynamics process

1) Force vector:

$$\begin{aligned} \mathbf{s} = K\mathbf{r} \\ \triangleq [s_1\ s_2\ \cdots\ s_i\ \cdots\ s_N]^T \end{aligned} \tag{4}$$

As a result of the stimulus interpretation module, $\mathbf{r}$ determines the position of the emotion in affect space. Vector $\mathbf{s}$ is the force that moves the emotion in the affect space, calculated by multiplying $\mathbf{r}$ by elasticity matrix K.

2) Emotion vector:

$$M\ddot{\mathbf{e}} + C\dot{\mathbf{e}} + K\mathbf{e} = \mathbf{s}$$
$$\mathbf{e} \triangleq [e_1\ e_2\ \cdots\ e_i\ \cdots\ e_N]^T \tag{5}$$

The second-order differential equation, emotion dynamics, is used to calculate vector $\mathbf{e}$ that represents the current emotion position of the robot in the N-dimensional affect space. The values of the inertia matrix M, viscosity matrix C, and elasticity matrix K control the characteristics such as speed and oscillation of the emotional movements in the coordinate system. The matrices M (inertia), C (viscosity), and K (elasticity) determine the dynamic characteristics of emotional transitions, such as speed, damping, and overshoot. Furthermore, these values are time-varying rather than constant, allowing for adaptive emotional responses. In addition to this variability, the inclusion of the matrix K ensures that the vector $\mathbf{e}$ converges back to its neutral state after a stimulus.

3) Dynamic parameters control: Variable emotional dynamics in the robot's expressions were modeled using a second-order mass-spring-damper system, where the system matrices are defined as $M = \mathrm{diag}(m_1, m_2, \ldots, m_N)$, $C = \mathrm{diag}(c_1, c_2, \ldots, c_N)$, and $K = \mathrm{diag}(k_1, k_2, \ldots, k_N)$. Among these parameters, only the viscosity matrix C was treated as a variable to enable dynamic modulation based on changes in emotional valence ($|\Delta x_1|$).

The rising time t_r and damping ratio ζ are computed using the following equations:

$$\ddot{\mathbf{e}} + 2\mathbf{Z}\boldsymbol{\Omega}_n\dot{\mathbf{e}} + \boldsymbol{\Omega}_n^2\mathbf{e} = \mathbf{s}$$
$$\zeta_i = \frac{c_i}{2\sqrt{m_i k_i}}, \quad \mathbf{Z} = \mathrm{diag}(\zeta_1, \zeta_2, \ldots, \zeta_N) \tag{6}$$
$$\omega_{n_i} = \sqrt{\frac{k_i}{m_i}}, \quad \boldsymbol{\Omega}_n = \mathrm{diag}(\omega_{n_1}, \omega_{n_2}, \ldots, \omega_{n_N})$$

With M and K held constant, t_r remained invariant ($t_{r_i} \approx 1.8/\omega_{n_i}$), allowing only ζ to vary. This isolated modulation of damping ratio enabled control over the degree of dynamic behavior while preserving the temporal structure of expressions. As the value of ζ decreases, the intensity of the transient overshoot and oscillation increases, thereby facilitating more exaggerated emotional expression.

C. Emotional expression generation

After the emotion vector $\mathbf{e}$ is determined through emotion dynamics, the robot generates corresponding expressions via its modalities.

$$\begin{aligned}\mathbf{p}_{(D_p\times 1)} &= T_{(D_p\times N)}\mathbf{e}_{(N\times 1)} \\ &\triangleq [p_1\ p_2\ \cdots\ p_i\ \cdots\ p_{D_p}]^T,\ -1 \le p_i \le 1 \\ {}^r\mathbf{p}_{(D_p\times 1)} &= f(\mathbf{p}_{(D_p\times 1)}), \qquad D_p \ge N\end{aligned} \tag{7}$$

A nonlinear function f is then applied to vector $\mathbf{p}$ to enforce modality-specific constraints, such as saturation at 1 or clipping at 0, ensuring that the resulting control vector ${}^r\mathbf{p}$ produces valid commands.

IV. ROBOT PROTOTYPE IMPLEMENTATION

As illustrated in Fig. 1 and the supplementary video, the prototype robot features a 13.3-inch OLED display and a built-in speaker for emotional expression. The system is controlled by a Raspberry Pi 5, integrating onboard sensors (IMU, camera, and touch) and two DC motors for physical movement.

A. Stimulus interpretation (demonstration only)

For demonstration purposes, we utilized two sensors—a touch sensor and a camera—to detect the user's emotional input. The touch sensor distinguishes between stroking (positive) and tapping (negative) interactions, while the camera captures facial expressions such as smiling (positive) and narrowing the eyes (negative). Note that the stimulus interpretation module was only used in the demonstration video and was not used in the user experiments.

1) Touch input: For touch-based affective input, the system computes the RMS acceleration of the 3-axis IMU signal over a 2-second sliding window ($n = 200$ samples, 0.01s interval) [27]. The valence contribution $x_{\text{touch}}(t)$ is then derived based on the ratio between $acc_{\text{rms},k}$ and a boundary value $acc_{\text{mid}} = 2g$, which differentiates hitting ($> acc_{\text{mid}}$) from stroking ($< acc_{\text{mid}}$). This formulation enables the robot to interpret high-magnitude events as negative and gentle touches as positive affective cues.

2) Face recognition: Facial valence cues were extracted using the MediaPipe BlendShape tool. Two coefficients—smile intensity and eye squinting—were obtained, and for each modality the higher value across the left and right sides was selected. These values were then compared against empirically determined baselines ($R_{\text{smile,baseline}} = 0.45, \quad R_{\text{squint,baseline}} = 0.6$).

If the smile ratio exceeded its baseline for at least three seconds, a positive valence increment V_{pos} was applied to the facial state $x_{\text{face}}(t)$. Conversely, sustained squinting without smiling applied a negative valence update V_{neg} to $x_{\text{face}}(t)$. Parameter values were set as follows: $V_{\text{pos}} = 0.25, \quad V_{\text{neg}} = -0.35$. This design prioritizes sensitivity to negative cues by assigning greater weight to squinting, thereby eliciting stronger corrective responses when potential discomfort or displeasure is detected.

3) Computation of force vector via valence-based fusion: In this prototype, we map all sensory inputs into a single, one-dimensional valence space, denoted by the subscript 1. The unified valence score $x_1(t)$ represents the aggregation of individual modality inputs (i.e., x_{touch} and x_{face}). The computation of $x_1(t)$ is defined piecewise to incorporate both stimulus accumulation and the gradual forgetting of past inputs. At the time points t_k, which represent the timestamps of touch or facial stimuli events, the valence score is updated by adding the current touch and face valence values to the previous accumulated value $x_1(t_k^-)$, where t_k^- denotes the moment just before the update at t_k. Also, since weights were not assigned to each input modality in this study, the w_i value was set to 1.

Between stimulus events, for $t_k < t < t_{k+1}$, the accumulated valence decays exponentially with a half-life $T_{1/2}$, representing the fading memory of previous stimuli over time. This exponential decay ensures that the robot's emotional state gradually returns toward neutrality in the absence of new stimuli, balancing responsiveness with emotional stability.

$$x_1(t) = \begin{cases} x_1(t_k^-) + x_{\text{touch}}(t) + x_{\text{face}}(t), & t = t_k \\ x_1(t_k) \times \left(\frac{1}{2}\right)^{\frac{t-t_k}{T_{1/2}}}, & t_k < t < t_{k+1} \end{cases} \tag{8}$$

The half-life $T_{1/2}$ was set to $T_{1/2} = 300$ seconds (5 minutes). This was chosen to minimize its influence during short interactions. In (1), the emotion amplitude a_i is presented as a function of the accumulated valence x_1 for the prototype robot. The parameters μ_i and σ_i define the centers and spreads of the Gaussian distributions corresponding to each of the six basic emotions, respectively. Specifically, $\mu_i = \{ -0.3, -0.6, -1.0, 1.0, -0.8, 0.5 \}$ and $\sigma_i = \{0.1, 0.1, 0.1, 0.2, 0.1, 0.2\}$ were empirically determined to reflect the emotional response characteristics of the system. In this model, the variable $x_1(t)$ denotes the current valence score, with the amplitude representing the degree to which the specific emotion is activated.

B. Emotion dynamics

The mass matrix M was fixed as the identity, while the diagonal elements of the elasticity matrix K were set to 200, ensuring a rise time of approximately 127 ms under all conditions. This configuration provides a response speed slightly faster than that of humans, while allowing controlled variation through C.

Five levels of expression intensity were generated using five predefined damping ratios for the experiment in Section V: $\zeta = [1.0, 0.7, 0.5, 0.3, 0.2]$.

The dynamic viscosity C was computed as:

$$\begin{aligned} c_i &= c_{\text{init}} \left(2 - \frac{1}{1 - \alpha_i |\Delta x_1|}\right), \quad |\Delta x_1| < 0.5 \\ \boldsymbol{\alpha} &= [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_i \ \cdots \ \alpha_N]^T, \quad -1 \le \alpha_i \le 1 \end{aligned} \tag{9}$$

The entry value of c_{init} is 28.3, which corresponds to a damping ratio ζ = 1.0 when $|\Delta x_1|$ = 0. This formulation causes the damping ratio ζ to decrease as the magnitude of the emotional change $|\Delta x_1|$ increases, thereby producing more expressive and overshooting behavior in the robot's motion. The coefficient vector $\boldsymbol{\alpha}$ is determined empirically through the user studies presented in this paper to align the robot's dynamic response with human perceptual characteristics. The specific derived values and their justification are detailed in Section VII-C as a key outcome of our experimental analysis.

C. Emotional expression generation

As shown in Fig. 2, the transformation matrix T maps vector $\mathbf{e}$ to CPs vector $\mathbf{p}$, where each modality has its own transformation, e.g., $T_{\text{face}}, T_{\text{motion}}, T_{\text{sound}}$. Unlike the facial and motion modalities that directly use $\mathbf{e}$, the sound modality derives its CPs from the stimulus vector $\mathbf{r}$, enabling immediate responsiveness. Sound playback is triggered only when $\mathbf{e}$ enters a defined emotional region U, at which point an open sound control signal initiates the appropriate cue. This mechanism ensures that auditory expressions remain both responsive to stimuli and synchronized with the robot's overall emotional state.

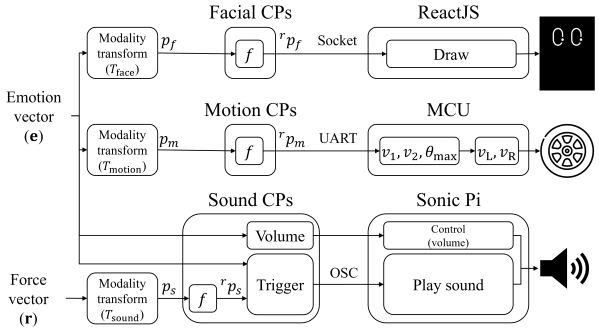

Fig. 2. Transformation of the emotion vector $\mathbf{e}$ into control points $\mathbf{p}$ via transformation T (T_{face}, T_{motion}, T_{sound}). This pipeline enables the robot to generate facial, motion, and sound expressions corresponding to the underlying emotional state.

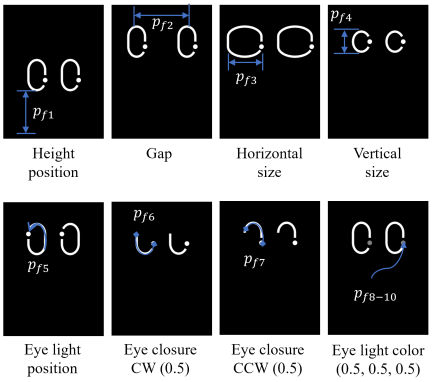

Fig. 3. Facial expressions when a single control point (CP) is active (1 or 0.5), while all other CPs are set to 0.

1) Facial expression: For emotional expressions through the facial modality, CPs were designed to represent key facial features, with each basic emotion defined by a specific CP configuration. The facial design adopts an iconic eye motif and employs 10 CPs in total (see Fig. 3, Table I). Although eye light color is not directly associated with facial AUs, it was included as a CP to enhance expressiveness. The neutral expression corresponds to all CP values set to zero. Each CP controls a specific aspect of the eye's shape, such as height, gap, width, and eye light position. The detailed transformation functions and parameter ranges for all 10 facial CPs were established based on our previous work [28] and are summarized in Table I and Fig. 3.

2) Motion expression: The robot's emotional expression through motion is modeled as a repetitive pattern defined by θ_{range} and two wheel speeds (v_1, v_2). The parameter θ_{range}

TABLE I

TRANSFORMATION FUNCTIONS f FOR EACH CONTROL POINT

Name	$f(p_i)$	**Range**
Height position	$(p_{f1}+1)\times 320\text{px}+160\text{px}$	480–800 px
Gap	$(p_{f2}+1)\times 80\text{px}$	80–160 px
Horizontal width size	$(p_{f3}+1)\times 160\text{px}$	160–320 px
Vertical height size	$(p_{f4}+1)\times 50\text{px}$	50–100 px
Eye light position	$(p_{f5}+1)\times 50\text{px}$	0–100%
Eye closure CW	$p_{f6}\times 100$	0–100%
Eye closure CCW	$p_{f7}\times 100$	0–100%
Eye light color (RGB)	$(1-p_{f8\text{–}10})\times 255$	255–0
Rotation range (θ_{range})	$p_{m1}\times 100^\circ$	0–100°
Motor speed (v_1)	$p_{m2}\times v_{\max}$	$[-v_{\max}, v_{\max}]$
Speed ratio (v_2)	$p_{m3}\times v_1$	$[-\lvert v_1\rvert, \lvert v_1\rvert]$
MIDI number (n_{MIDI})	$p_{s1}\times 100$	0–100
Attack (t_{att})	$p_{s2}\times 2s$	0–2 seconds
Attack level (l_{att})	$p_{s3}\times 4$	0–4
Decay (t_{dec})	$p_{s4}\times 2s$	0–2 seconds
Decay level (l_{dec})	$p_{s5}\times 4$	0–4
Sustain (t_{sus})	$p_{s6}\times 2s$	0–2 seconds
Sustain level (l_{sus})	$p_{s7}\times 4$	0–4
Release (t_{rel})	$p_{s8}\times 2s$	0–2 seconds
Divisor (D)	$p_{s9}\times 1000$	0–1000

limits the robot's rotation range, while v_1 and v_2 denote the left and right wheel speeds ($|v_2| \leq |v_1|$). When the yaw angle θ reaches θ_{range}, the wheel speeds are swapped, producing the motion pattern. To control this behavior, three motion CPs (p_m) are defined to regulate the rotation range and wheel speeds. The exact transformation functions and parameter ranges were established based on our previous work [29] and are detailed in Table I.

3) Sound expression: This work focuses on expressing a robot's emotions through non-verbal sound effects rather than speech, aiming for a character style similar to Wall-E, where emotions are conveyed via tone, pitch, and modulation. To achieve this, we employed frequency modulation (FM) synthesis using Sonic Pi [1], which enables real-time programmable sound generation.

The synthesized sound is defined as:

$$\begin{aligned} f_c &= 440\times 2^{\frac{n_{\text{MIDI}}-69}{12}} \\ y(t) &= E(t)\cdot\sin\left(2\pi f_c t + I\sin\left(2\pi\tfrac{f_c}{D}t\right)\right) \end{aligned} \tag{10}$$

where $E(t)$ denotes the amplitude envelope shaped by Attack-Decay-Sustain-Release (ADSR) parameters. The modulation index I controls the depth of timbral variation (fixed at $I = 1$ in this work), while the divisor D determines the harmonic structure by scaling the modulating frequency relative to the carrier.

This design allows auditory cues to be pre-generated, parametrically modulated, and promptly triggered in synchrony with the robot's affective state, ensuring responsive and expressive sound-based emotion rendering.

All envelope parameters p_{s1}–p_{s9}, which control the pitch, ADSR phases, and output levels, are normalized within the range $[0, 1]$. Their specific transformation functions and detailed parameter mappings are summarized in Table I.

[1]https://sonic-pi.net/

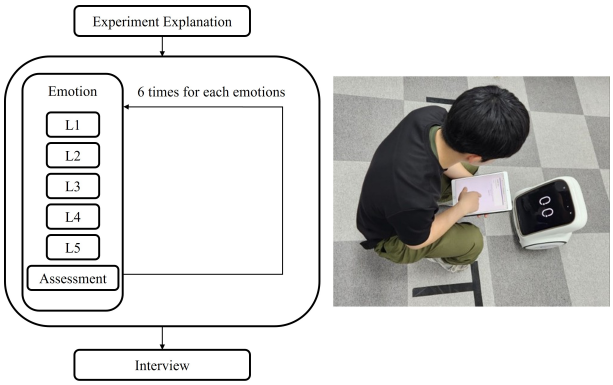

Fig. 4. Left: Flowchart of the experiment for evaluating liveliness and naturalness across five dynamic levels (L1–L5). Right: Experimental setup showing the environment in which participants observed and rated the robot's emotional expressions.

4) User Survey-Based Determination of Control Points: Table II summarizes the CP values determined for each emotion across different modalities. These CP values were not arbitrarily chosen; rather, they were derived from a previous study [1] in which user surveys were conducted. Specifically, multiple candidate values were evaluated, and the ones that received the highest user preference scores were selected as the final CP values. Therefore, the CP values adopted in this study reflect both subjective reliability and actual user experience, providing a reasonable basis that accounts for emotional characteristics and modality-specific responses.

V. EXPERIMENT

The user study was conducted with 22 participants (14 male, 8 female; age range 22–42 years, M = 26.9, SD = 4.75) from diverse academic backgrounds. The study examined how variations in dynamic parameters influence the perception of emotional expression across six basic emotions: anger, disgust, fear, happiness, sadness, and surprise. As shown in Fig. 4, participants observed the robot's expressions across five dynamic levels (L1—L5), where lower damping ratios produced more exaggerated and expressive behaviors across facial, motion, and sound channels.

To ensure participants could establish a stable internal baseline for comparing subtle kinematic differences, the presentation order of the six emotions and their respective five dynamic levels was fixed rather than randomized. This structured sequence allowed for a systematic evaluation of increasing dynamic intensity within each emotional category. Evaluation was conducted in a controlled laboratory setting, with participants rating each expression on 7-point scales for liveliness (to evaluate vivacity and clarity) and naturalness (to assess reliability and appropriateness) [19], [30].

Q1. *How lively did the robot's emotional expressions feel?* (1 = very low, 7 = very high)

Q2. *How natural was the expression of this emotion?* (1 = very unnatural, 7 = very natural)

In addition to quantitative ratings, participants provided qualitative feedback through open-ended questions about

TABLE II

CONTROL POINT (CP) VALUES FOR FACIAL, MOTION, AND SOUND EXPRESSIONS ACROSS SIX BASIC EMOTIONS

	Facial (T_{face})										Motion (T_{motion})			Sound (T_{sound})							
	p_{f1}	p_{f2}	p_{f3}	p_{f4}	p_{f5}	p_{f6}	p_{f7}	p_{f8}	p_{f9}	p_{f10}	p_{m1}	p_{m2}	p_{m3}	p_{s1}	p_{s2}	p_{s3}	p_{s4}	p_{s5}	p_{s6}	p_{s7}	p_{s8}	p_{s9}
Anger (e_1)	0.50	0.00	0.59	0.00	0.40	0.65	0.25	0.00	1.00	1.00	0.00	0.70	1.00	0.81	0.03	0.75	0.10	0.63	0.15	0.38	0.30	0.08
Disgust (e_2)	−0.06	1.00	−1.00	−0.26	0.60	0.32	0.47	0.00	0.00	0.00	0.00	−0.40	1.00	0.60	0.03	0.20	0.10	0.18	0.20	0.15	0.30	0.01
Fear (e_3)	1.00	−0.46	0.02	−1.00	0.29	0.00	0.00	1.00	1.00	1.00	0.10	−0.60	−0.30	0.83	0.25	0.08	0.25	0.13	0.25	0.08	0.55	0.06
Happiness (e_4)	0.32	0.04	0.21	1.00	0.36	0.00	0.50	0.00	0.00	0.00	0.30	−0.99	−1.00	0.70	0.01	0.38	0.35	0.30	0.20	0.38	0.25	0.07
Sadness (e_5)	1.00	−0.46	1.00	1.00	0.50	0.48	0.00	0.67	0.27	0.08	0.10	−0.20	0.60	0.62	0.15	0.10	0.15	0.05	0.15	0.05	1.00	1.00
Surprise (e_6)	−0.02	−0.21	−0.52	−1.00	−0.44	0.00	0.00	0.05	0.00	1.00	0.00	−0.90	1.00	0.85	0.15	0.45	0.00	0.18	0.00	0.13	0.25	0.10

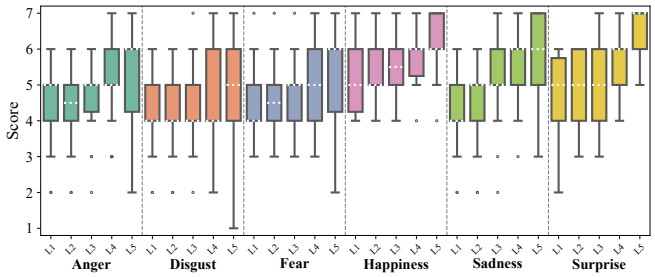

Fig. 5. Participant ratings of perceived liveliness for each emotion across five dynamic levels.

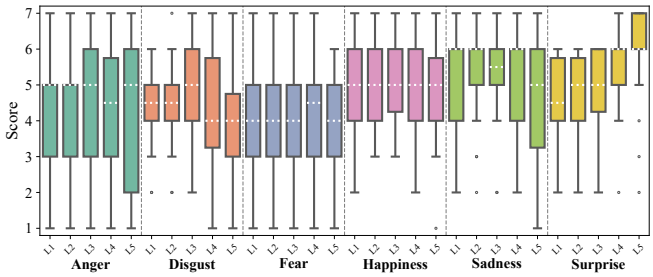

Fig. 6. Participant ratings of perceived naturalness for each emotion across five dynamic levels.

their impressions of the varying dynamic levels, the emotion-specific appropriateness of the expressions, and potential differences across modalities.

VI. RESULT

A. Perceived liveliness across dynamic levels

To evaluate the impact of damping-based dynamic modulation, a Friedman test was conducted on the liveliness ratings for each of the six emotions across five dynamic levels. The test revealed a statistically significant effect of dynamic levels on perceived liveliness for all emotional categories. The most prominent effects were observed in surprise ($\chi^2(4) = 50.3, p < .001, W = 0.57$) and sadness ($\chi^2(4) = 30.7, p < .001, W = 0.35$), both demonstrating substantial effect sizes (W). Happiness ($\chi^2(4) = 26.6, p < .001, W = 0.30$) and anger ($\chi^2(4) = 22.6, p < .001, W = 0.26$) also showed moderate effect sizes with high significance. While disgust ($\chi^2(4) = 14.6, p = .006, W = 0.17$) and fear ($\chi^2(4) = 13.0, p = .011, W = 0.15$) exhibited relatively smaller effect sizes, they remained statistically significant, confirming that varying the damping ratio consistently alters the perceived intensity of the robot's multimodal expressions.

Post-hoc pairwise comparisons were performed using the Durbin-Conover test, which inherently provides adjusted p-values to control for multiple comparisons (see Fig. 5). In surprise, while almost all pairs exhibited significant differences ($p < .001$), no significant differences were observed among the lower dynamic levels (L1, L2, L3). Specifically, unlike other emotions, a significant difference between L4 and L5 was observed only in the surprise emotion ($p = .001$). For sadness, happiness and anger, significant increases in liveliness were found when comparing the lower levels (L1, L2) to the more dynamic levels (L4, L5) such as L2–L4 (sadness: $p < .001$, happiness: $p = .003$, anger: $p < .001$). However, for disgust and fear, significant differences only began to emerge at more extreme levels, such as L1–L5 (disgust: $p = .002$, fear: $p = .001$). These results suggest that increasing dynamic expression intensity generally enhances perceived liveliness, though the degree of sensitivity varies by emotion.

B. Perceived naturalness across dynamic levels

As illustrated in Fig. 6, the Friedman test revealed that dynamic modulation significantly impacted perceived naturalness only for surprise ($\chi^2(4) = 30.7, p < .001, W = 0.51$) and disgust ($\chi^2(4) = 9.78, p = .044, W = 0.16$). Post-hoc Durbin-Conover tests for surprise showed no significant differences among the lower levels (L1–L2, $p > .05$), but revealed increases in naturalness when comparing lower levels (L1, L2) to higher levels (L4, L5; $p < .001$). Notably, a significant difference was maintained even between L4 and L5 ($p = .048$). For disgust, post-hoc tests revealed that naturalness significantly decreased between L3 and L5 ($p = .002$). This suggests that while intermediate dynamic levels (L3) are perceived as most natural for disgust, excessive exaggeration at L5 leads to a significant drop in naturalness.

Conversely, for the remaining four emotions (anger, fear, happiness, sadness), although no statistically significant effects were found, the average ratings peaked at L3 (anger: 4.55, fear: 4.32, happiness: 5.18, sadness: 5.27) and began to decline from L4 (anger: 4.36, fear: 4.18, happiness: 5.00, sadness: 5.09).

C. Perceived differences between weak and strong expressions

To understand user preferences for conveying different emotional intensities, participants were asked to select the

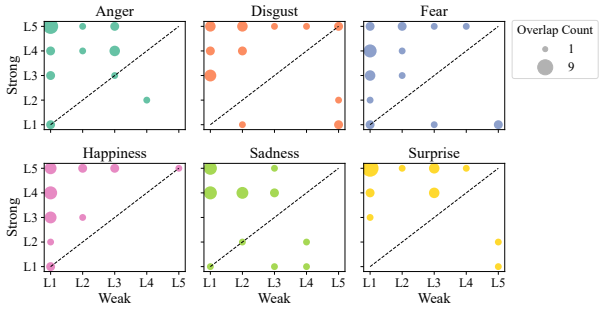

Fig. 7. Scatter plots illustrating the specific dynamic levels (L1–L5) selected by participants as most suitable for expressing weak (x-axis) versus strong (y-axis) emotional intensities. Each point represents a (weak, strong) score pair per participant.

most appropriate dynamic level (L1 to L5) for expressing a "weak" and a "strong" version of each emotion. Fig. 7 presents scatter plots comparing individual participant ratings for emotional expressions at weak versus strong stimulus intensities across six basic emotions. Each data point corresponds to a participant's paired selections, with the x-axis representing the weak choice and the y-axis representing the strong choice. The plots can be divided by the diagonal into three groups: above the line (stronger perceived at higher level), on the line (no difference), and below the line (stronger perceived at lower level). Most participants are distributed above the dotted line, with only a few participants either on the dotted line or below it.

VII. DISCUSSION

A. Effect of increasing dynamic level on perceived liveliness

Results indicate that higher levels of dynamic expression were generally perceived as more lively. Liveliness ratings showed a clear upward trend across the five dynamic stages, with the strongest contrasts emerging between lower (L1, L2) and higher (L4, L5) levels. Several participants noted that the differences between L1 and L2 were subtle, while higher levels were perceived as distinctly more energetic and expressive. For example, P13 remarked that "the higher the level for happiness and anger, the more lively and natural it felt," and P12 observed that "anger expressions were more convincing at level 5."

The emotion surprise consistently benefited from increased dynamics, with ratings rising robustly and linearly across participants. This suggests that surprise is particularly well-suited to exaggerated and high-energy expressions. By contrast, emotions such as anger displayed more complex patterns: moderate intensities were often rated positively, but the highest level (L5) was sometimes described as "too dramatic" or "unnatural."

Overall, increasing expression intensity enhances liveliness, but the effect is not uniform across emotions or individuals. Some emotions appear to have a perceptual threshold beyond which additional dynamism reduces believability and compromises naturalness.

B. Variability in naturalness perception across emotions and individuals

While liveliness generally increased with expression intensity, perceived naturalness followed a more nuanced pattern. Participants reported distinct dynamic levels that maximized perceived naturalness for each emotion: subtle dynamics yielded the highest naturalness ratings for sadness, disgust, and fear, whereas stronger dynamics were evaluated as more natural for surprise and happiness. This highlights that naturalness depends not only on movement magnitude but also on its contextual appropriateness.

Individual perception of these dynamic expressions varied considerably. While consensus was reached regarding surprise, participants' assessments of the most natural level of dynamism for other emotions differed. This finding indicates that individuals' sensitivity to a robot's emotional expression is influenced by their interpretive tendencies, particularly at higher levels of dynamism.

Another recurring observation was cross-modal inconsistency. When facial expressions appeared more exaggerated than body motions (or vice versa), participants perceived reduced naturalness. These findings imply that uniform modulation strategies may be suboptimal. Instead, tailoring expression dynamics to both emotion type and individual preference could foster more believable and coherent affective displays.

C. Selection of dynamic level according to emotional intensity

Overall, results suggest that stronger emotions are associated with higher levels of dynamic movement, while weaker emotions are better conveyed through lower dynamics. Participants rated high dynamic levels as more appropriate for intense emotions such as anger and surprise, and lower levels for subdued emotions such as sadness or disgust. This supports the hypothesis that emotional intensity can be modulated through kinematic parameters such as K and velocity.

Contrary to expectations, Fig. 7 shows that some participant groups rated low dynamism as more suitable for strong negative emotions like fear or sadness, reporting that restrained movements felt more intense and authentic than exaggerated ones.

Taken together, these findings indicate that while high dynamics are often interpreted as markers of emotional intensity, context and emotion type critically shape users' perceived feelings. This underscores the importance of tuning expression parameters not only to emotion strength but also to individual perceptual preferences.

Based on these results, the α value for the prototype robot in (9) was set as follows: $\alpha = [0.823, 0.888, 0.823, 0.888, 0.823, 0.888]$. Adjusting this value controls the degree of dynamic expression when conveying strong intensity for each emotion, and these parameter settings were also reflected in the demonstration video to showcase the resulting expressive behaviors.

VIII. CONCLUSION

This work examined how dynamic levels of emotional expression influence the perception of a social robot's affective displays. Results showed that participants could reliably distinguish between most dynamic levels, with the exception of the lowest two levels (L1, L2). While liveliness ratings generally increased with higher dynamics, naturalness ratings revealed a different pattern: most emotions peaked at moderate intensity levels, suggesting a perceptual threshold beyond which expressions appear exaggerated or unnatural. Some participants perceived stronger dynamics as weaker, indicating the presence of emotion- and user-specific sweet spots in expression intensity.

These individual differences highlight the need for personalized expression adjustments, yet our findings still provide fundamental baseline guidelines for robot design. Due to the common trade-off between liveliness and naturalness at high intensities, simultaneous optimization is inherently challenging. Designers must generally prioritize based on emotional context: highly arousing emotions (e.g., surprise) can be expressed energetically to convey energy, while subtle or negative emotions (e.g., sadness, fear) require prioritizing naturalness to maintain credibility and avoid artificiality.

One limitation is that our findings, based on a specific robot platform, may not fully generalize to robots with different morphologies. Future research should investigate modality-specific effects by testing each channel individually, providing tailored guidance for adaptive emotion delivery. Additionally, the ecological validity of our study is limited by the controlled laboratory setting. While this setup was essential to isolate the perceptual effects of specific dynamic parameters (e.g., damping ratio), participants' explicit awareness of the task may have heightened their sensitivity compared to spontaneous "in-the-wild" interactions. Future work should evaluate these expressions during unscripted, task-oriented interactions to verify their effectiveness in more ecologically valid environments.

REFERENCES

[1] H. Park, "Emotion engine with dynamic characteristic changes for multimodal emotion expression in social robots," Ph.D. dissertation, Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea, 2025.

[2] C. L. Breazeal, *Designing sociable robots*. MIT press, 2002.

[3] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics Auton. Syst.*, vol. 42, no. 3-4, pp. 143–166, 3 2003.

[4] S. Ojha, J. Vitale, and M.-A. Williams, "Computational emotion models: a thematic review," *Int. J. Soc. Robotics*, vol. 13, no. 6, pp. 1253–1279, 9 2021.

[5] A. Sripathy, A. Bobu, Z. Li, K. Sreenath, D. S. Brown, and A. D. Dragan, "Teaching robots to span the space of functional expressive motion," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 10 2022, pp. 13 406–13 413.

[6] P. Alves-Oliveira, K. Mihata, R. Karim, E. A. Bjorling, and M. Cakmak, "Flex-sdk: An open-source software development kit for creating social robots," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, 10 2022, pp. 1–10.

[7] V. N. Antony, M. Stiber, and C.-M. Huang, "Xpress: A system for dynamic, context-aware robot facial expressions using language models," in *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 3 2025, pp. 958–967.

[8] P. Ekman, "Are there basic emotions?" *Psychol. Rev.*, vol. 99, pp. 550–553, 1992.

[9] C. Breazeal, "Function meets style: Insights from emotion theory applied to hri," *IEEE Trans. on Syst. Man Cybern. Part C: Appl. Rev.*, vol. 34, pp. 187–194, 5 2004.

[10] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Res. Pers.*, vol. 11, no. 3, pp. 273–294, 9 1977.

[11] J. A. Russell, "A circumplex model of affect." *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 12 1980.

[12] R. Plutchik, *The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice*. American scientist, 2001, vol. 89, no. 4.

[13] B. Gonsior, S. Sosnowski, M. Buss, D. Wollherr, and K. Kuhnlenz, "An emotional adaption approach to increase helpfulness towards a robot," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 10 2012, pp. 2429–2436.

[14] M. Zheng, Y. She, F. Liu, J. Chen, Y. Shu, and J. XiaHou, "BabeBay-A companion robot for children based on multimodal affective computing," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 3 2019, pp. 604–605.

[15] F. Yan, X. Yang, N. Li, X. Yu, and H. Zhai, "Emotion generation and transition of companion robots based on plutchiks model and quantum circuit schemes," *Secur. Commun. Networks*, vol. 2021, pp. 1–15, 8 2021.

[16] H. Miwa, T. Okuchi, K. Itoh, H. Takanobu, and A. Takanishi, "A new mental model for humanoid robots for human friendly communication introduction of learning system, mood vector and second order equations of emotion," in *2003 IEEE International Conference on Robotics and Automation*, vol. 3. Taipei, Taiwan: IEEE, 2003, pp. 3588–3593.

[17] H. S. Lee, J. W. Park, and M. J. Chung, "A linear affectexpression space model and control points for mascot-type facial robots," *IEEE Trans. on Robotics*, vol. 23, no. 5, pp. 863–873, 10 2007.

[18] C. Breazeal, "Emotion and sociable humanoid robots," *Int. J. Human-Computer Stud.*, vol. 59, no. 1-2, pp. 119–155, 7 2003.

[19] M. Saerbeck and C. Bartneck, "Perception of affect elicited by robot motion," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 3 2010, pp. 53–60.

[20] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 3 2013, pp. 301–308.

[21] G. Hoffman and W. Ju, "Designing robots with movement in mind," *J. Human-Robot Interact.*, vol. 3, no. 1, p. 89, 3 2014.

[22] M. J. Gielniak and A. L. Thomaz, "Spatiotemporal correspondence as a metric for human-like robot motion," in *Proceedings of the 6th international conference on Human-robot interaction*. New York, NY, USA: ACM, 3 2011, pp. 77–84.

[23] P. Ekman and W. V. Friesen, "Facial action coding system," *Environ. Psychol. & Nonverbal Behav.*, 1978.

[24] K. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Commun.*, vol. 40, no. 1-2, pp. 227–256, 4 2003.

[25] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behav. Brain Sci.*, vol. 31, no. 5, pp. 559–575, 10 2008.

[26] H. Park, J. Lee, T. Dzhoroev, B. Kim, and H. S. Lee, "Expanded linear dynamic affect-expression model for lingering emotional expression in social robot," *Intell. Serv. Robotics*, vol. 16, no. 5, pp. 619–631, 11 2023.

[27] H. Park, J. Lee, and H. S. Lee, "Adaptive emotional expression in social robots: A multimodal approach to dynamic emotion modeling," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. Atlanta, USA: IEEE, 5 2025, pp. 5504–5511.

[28] H. Kim, H. Park, S. J. Hwang, and H. S. Lee, "A control point based facial expression for smooth facial display in social robot expression transitions," in *2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 8 2025.

[29] H. Park, J. Lee, T. Dzhoroev, B. Kim, and H.-S. Lee, "Developing a dynamic expression model that can simultaneously control robot's facial and movement expressions," *J. Inst. Control. Robotics Syst.*, vol. 30, no. 1, pp. 8–12, 1 2024.

[30] C. Bartneck, D. Kuli, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *Int. J. Soc. Robotics*, vol. 1, pp. 71–81, 1 2009.