Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ30075

Extending Emotional Well-Being in Digital Art Therapy: The Role of Dynamic Audiovisual Feedback

Chen Peng[1], Jiatong Ma[1*], Liqiao Zhong[2]

[1]*Queen Mary School Hainan, Beijing University of Posts and Telecommunications, Lingshui, China*
[2]*No.2 High School of East China Normal University, Shanghai, China*
**Corresponding Author. Email: majt181@bupt.edu.cn*

Abstract. In this study, our team focused on addressing the stress-related emotional needs of urban students—specifically ten participants aged 17 to 21—by developing and evaluating a keyboard-controlled interactive installation. This installation integrates natural visual elements and therapeutic sounds, with the overarching goal of enhancing emotional well-being within the context of digital art therapy. A within-subjects experiment compared three conditions: Multimodal Audiovisual Feedback, Monomodal Visual Feedback, and Passive Nature Exposure. Emotional change was measured with the PANAS scale, and engagement through interaction counts and total duration. Results revealed that the multimodal feedback condition significantly reduced Distressed scores ($t(9)=4.82$, $p=0.001$) and increased Attentive scores ($t(9)=4.97$, $p=0.0007$). Engagement was also higher (mean 1086.6 interactions, $p=0.039$). Gender and prior piano experience influenced outcomes, suggesting nuanced user responses. These results not only highlight the potential of multimodal interaction for rapid mood regulation but also lead our team to propose further investigation into how this approach can be applied more broadly in therapeutic settings.

Keywords: Human-Computer Interaction (HCI), Digital Art Therapy, Multimodal Audiovisual Feedback, Emotional Recovery in Digital Contexts, Keyboard-Controlled Interactive Installation, PANAS Scale, Behavioral Engagement Metrics

1. Introduction

College students routinely face stress in urban environments. Academic workload, social pressures, and the constant stimuli of city life can lead to chronic emotional strain. Short, restorative digital experiences could help relieve tension without demanding long time or resource investment. We propose a human–computer-interaction-based approach to embed therapeutic experiences in everyday digital contexts, allowing quick emotional "resets."

Multisensory installations have become increasingly prevalent in HCI as interactive environments that stimulate multiple senses—vision, hearing, and sometimes touch—to immerse users in responsive experiences. Such systems have been shown to elicit affective responses and promote psychological well-being [1, 2]. They can deliver the restorative effects of nature while granting users active agency rather than passive observation. Building on positive-psychology

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ30075

theory that small moments of joy can accumulate into emotional resilience [3], we aim to design dynamic systems that create quick, engaging pathways to emotional repair.

To address this, we developed an interactive digital installation using keyboard and mouse input to immerse users in a biophilic virtual environment with synchronized multisensory feedback. The project situates itself at the intersection of affective HCI, art therapy, and emotion regulation.

2. Related work

In recent years, growing interest has emerged in using digital media to support emotional health. This trend has been driven by advances in both technology and psychological science, encouraging researchers to explore how interactive systems can relieve stress and improve mood. Human beings naturally rely on sight, sound, and touch to regulate their emotions, and digital systems that engage several of these senses tend to generate stronger effects than those limited to a single modality.

2.1. Emotion regulation via physiological feedback

Physiological feedback—providing users with information about signals such as heart rate, skin conductance, or breathing—has long been recognized as an effective means of emotion regulation. Recent systematic reviews confirm that biofeedback systems can reliably enhance emotional control, although their effectiveness depends on the modality and immediacy of the feedback [4]. In particular, real-time feedback based on heart-rate variability or electrodermal activity helps users translate internal states into actionable cues, which in turn fosters self-awareness and calm. This kind of automatic responsiveness makes physiological feedback suitable for individuals with varying cognitive resources or technological familiarity, extending its reach in therapeutic settings.

2.2. Multimodal interaction & affective VR interfaces

Combining visual, auditory, and bodily signals into a single experience has become essential for emotion-aware systems. Controlled virtual-reality platforms such as the Magic XRoom have shown how adaptive tasks can be tuned according to users' skill and emotional state, creating reliable data for affect induction and recognition [5]. Such work illustrates the promise of closed-loop systems in which affect measurement and modulation occur simultaneously—a principle we sought to translate into our art-therapy context.

2.3. HCI-focused emotion regulation design

Human-Computer Interaction research on emotion regulation increasingly adopts theory-driven, user-centered design frameworks. Recent discussions emphasize deliberate choices in mapping psychological targets to interaction mechanisms. Techniques like affective haptics—touch-based cues that elicit or moderate emotion—are also entering immersive media to improve communication and presence. Meanwhile, interactive artworks such as Lozano-Hemmer's Pulse Room and teamLab's large-scale immersive spaces demonstrate how sensor-driven environments can transform bodily signals into collective, emotionally resonant experiences [6, 7]. These precedents reinforce the view that technological mediation can amplify empathy and connection when designed with affect in mind.

DOI: 10.54254/2755-2721/2026.TJ30075

2.4. Summary of gaps & positioning

Building upon this landscape, existing systems often focus on either single-modal, open-loop interventions or passive emotional exposure. What remains comparatively rare is the integration of active multimodal (visual + auditory + physiological) engagement within immersive environments, real-time closed-loop regulation using physiological signals, and adaptive interaction mechanisms informed by HCI frameworks.

Our proposed work addresses these gaps by designing a system that combines user-controlled multimodal input (e.g., keyboard or motion interactions) with real-time physiological feedback embedded in a VR context. The architecture emphasizes immediacy, explainability, and replicability, consistent with HCI-informed intervention design, and aims to measure emotion regulation effectiveness dynamically through engagement metrics, duration, and immediate affective trajectories.

3. Design

The objective of this study was to design and implement an interactive digital art therapy device that investigates the impact of three different experiential modes—multimodal feedback, single-modal (visual-only) feedback, and passive natural exposure—on users' emotional well-being. The device was developed using the Processing programming environment, enabling the seamless integration of visual and auditory feedback mechanisms inspired by natural motifs, such as butterflies, forests, and ambient nature sounds. These design choices were intended to evoke a sense of calmness and psychological restoration, aligning with the broader therapeutic aims of digital art-based emotional regulation.

In parallel with the Processing-based implementation, a high-fidelity interactive prototype was developed using Figma to simulate a potential mobile application version of the device [see Figure 1 and Figure 2]. This prototype preserved the visual aesthetics and interaction logic of the Processing version, ensuring conceptual and experiential consistency between platforms. The dual-prototype approach allowed for both empirical testing and user experience evaluation, facilitating future scalability and adaptation to mobile interfaces.

The device supports three experimental conditions [Table 1]:

Table 1. Three experimental conditions

Mode	Condition	Description
Mode 1	Negative control	Users gaze at a static forest background with no interactions or dynamic elements.
Mode 2	Single-Modal Visual	Pressing 'A' triggers random butterfly movements; arrow keys adjust position, no sound.
Mode 3	Multimodal	Keyboard inputs trigger butterfly movements and piano tones; arrow keys for tweaks.

DOI: 10.54254/2755-2721/2026.TJ30075



Figure 1. Welcome page

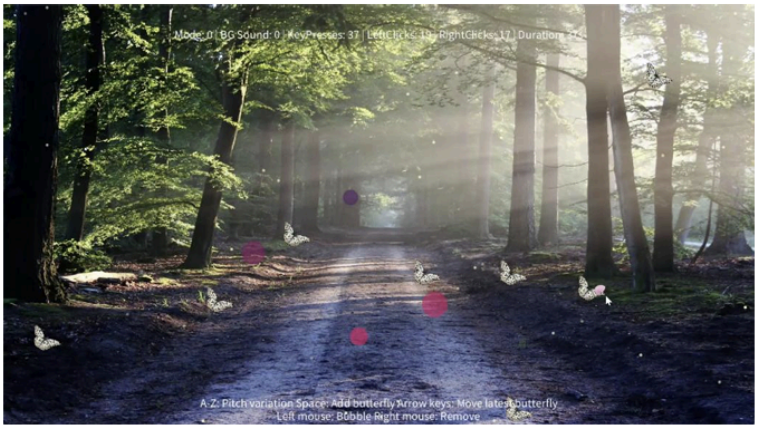


Figure 2. Main interface

User interaction with the device was primarily facilitated through keyboard and mouse inputs, enabling an intuitive yet expressive engagement with the digital environment. Specifically, the spacebar key was programmed to increase the number of butterflies displayed on screen while simultaneously triggering birdsong sound effects, thereby linking visual proliferation with auditory enrichment. The alphabet keys produced piano-like acoustic feedback and generated corresponding butterfly motion responses, creating a dynamic interplay between user input and environmental transformation. Furthermore, the arrow keys allowed participants to manipulate the position of a butterfly, simulating a sense of interactive spatial control and fostering immersion. Mouse inputs were also integrated into the interactive system: pressing the left mouse button resulted in an increase in the number of dynamic bubbles on the screen, whereas pressing the right mouse button reduced their number, providing users with tangible control over the system's visual density and rhythmic flow.Together these controls created a coherent sense of play, embodiment, and agency consistent with theories of restorative engagement [2, 3].

Through this multimodal design framework, the system sought to cultivate a soothing, participatory digital art experience, encouraging users to engage in gentle, exploratory interactions that could elicit positive emotional shifts and support emotional self-regulation.

DOI: 10.54254/2755-2721/2026.TJ30075

4. Methods

This study recruited a total of ten student participants, aged between 17 and 21 years, all of whom self-reported experiencing negative emotional states—such as stress, sadness, or related affective disturbances—within the preceding two weeks. The research adopted a within-subjects experimental design, enabling each participant to serve as their own control across multiple experimental conditions. Specifically, participants were exposed to three distinct intervention conditions, each designed to promote emotional well-being and relaxation: Negative Control (Natural Calm), Monomodal Visual Feedback (Monomodal Calm), and Multimodal Audiovisual Feedback (Multimodal Calm). This design facilitated direct comparison of individual responses across varying levels of sensory engagement. To control for potential order effects and reduce bias, the presentation sequence of the three conditions was randomized for each participant.

All experimental sessions were conducted in a quiet, controlled laboratory environment, carefully maintained to ensure a consistent setting and to minimize the influence of external distractions or environmental variability. The experimental procedure aimed to foster a neutral baseline atmosphere, allowing observed emotional changes to be attributed primarily to the intervention conditions rather than extraneous factors.

To evaluate participants' emotional states, the study utilized a modified version of the Positive and Negative Affect Schedule (PANAS) questionnaire [8]. This adapted instrument was designed to assess six discrete emotional dimensions, comprising four negative emotions—Distressed, Upset, Irritable, and Nervous—and two positive emotions—Attentive and Active. Participants were instructed to rate the intensity of each emotional state on a 10-point interval scale, ranging from 10 (indicating minimal intensity) to 50 (representing extreme intensity), with fixed increments of 10 (e.g., 20, 30, 40). These ratings reflected the subjective strength of each emotion as experienced during or immediately following each experimental condition, thereby providing a quantifiable measure of affective change associated with the different feedback modalities.

The experimental procedure for each session followed a structured process, as outlined below and illustrated in [Figure 3]:

DOI: 10.54254/2755-2721/2026.TJ30075

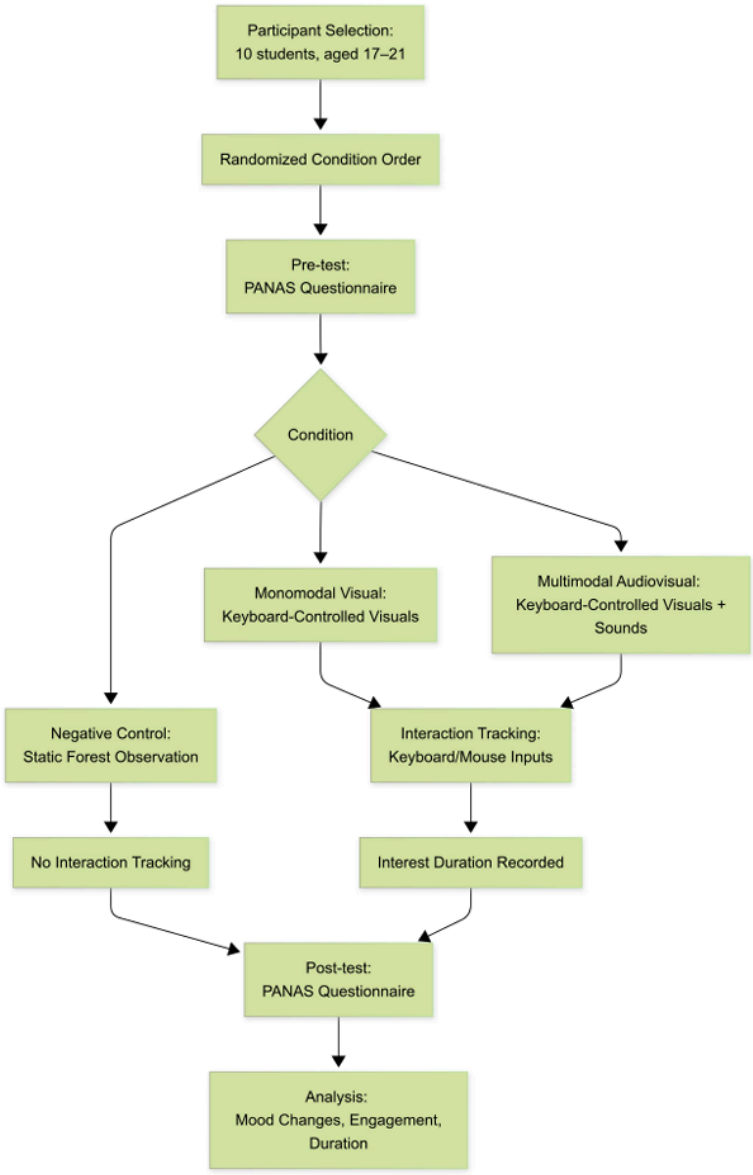


Figure 3. Experimental process

Pre-test Assessment: Participants completed the PANAS questionnaire to establish baseline emotional states prior to engaging with the assigned condition.

Regulation Phase: Participants interacted with one of the three conditions:

Negative Control (Natural Calm): Participants observed a static forest background without interactive elements or tracking.

Monomodal Visual Feedback (Monomodal Calm): Participants used keyboard inputs to trigger visual stimuli (e.g., butterfly movements) without auditory feedback.

Multimodal Audiovisual Feedback (Multimodal Calm): Participants used keyboard inputs to trigger both visual (e.g., butterfly movements) and auditory (e.g., piano tones, birdsong) stimuli. Interaction data, including keyboard presses and mouse clicks, were recorded to measure engagement. There was no time limit; participants ceased interaction when they felt the condition no longer provided emotional benefit, and this duration was recorded as "interest duration."

DOI: 10.54254/2755-2721/2026.TJ30075

Post-test Assessment: Immediately following the regulation phase, participants completed the PANAS questionnaire again to capture changes in emotional states.

To evaluate the effectiveness of each condition, three key metrics were analyzed:

Mood Changes: Differences between pre- and post-session PANAS scores were calculated to assess shifts in positive and negative emotions.

Engagement: For Monomodal and Multimodal conditions, the total number of keyboard presses and mouse clicks was recorded to quantify interaction levels.

Duration of Interest: The time spent engaging with each condition was measured to determine sustained attention.

This combination allowed both emotional and behavioral dimensions to be captured, a common practice in affective HCI research [1, 4].

Statistical analyses were performed using SPSS 27.0. A repeated-measures ANOVA compared the effects of feedback modality on positive and negative affect, while t-tests explored pairwise differences. Gender and piano experience were included as between-subject factors. Correlations were computed between interaction counts and affective change.

5. Results

This section presents the findings from the within-subjects experiment with 10 students (aged 17–21) evaluating three conditions—Negative control (Model 1), Monomodal Visual Feedback (Model 2), and Multimodal Audiovisual Feedback (Model 3)—for their efficacy in accelerating emotional recovery and sustaining engagement. To quantify the effect of our interactive digital art therapy device on emotional healing, we measured mood changes via the Positive and Negative Affect Schedule (PANAS), interaction counts (keystrokes and mouse clicks), duration of interest, and qualitative feedback. Subgroup analyses explore differences by gender and prior piano experience. Statistical tests used $\alpha = 0.05$, with paired t-tests for within-model comparisons and ANOVA with Tukey HSD post-hoc tests for between-model comparisons. We set $\alpha = 0.05$ as a conventional threshold to control the Type I error rate, a practical choice for our small sample that balances detecting real benefits with avoiding overinterpretation. Paired t-tests were chosen for within-model comparisons to compare each participant's pre- and post-session scores, accounting for individual differences and boosting reliability in our design. For between-model comparisons, ANOVA assessed overall differences across the three conditions, with Tukey HSD post-hoc tests pinpointing specific pairwise effects while managing the risk of false positives from multiple tests—though α could be adjusted (e.g., to 0.01) if stricter control were needed given our sample size.

5.1. Emotional recovery effects

5.1.1. Within-model comparisons (before vs. after)

Paired t-tests compared pre- and post-session PANAS scores for four negative emotions (Distressed, Upset, Irritable, Nervous) and two positive emotions (Attentive, Active) across the three models. [Table 2] summarizes the key findings, focusing on the most representative emotional shifts to highlight the models' effectiveness in recovery.

DOI: 10.54254/2755-2721/2026.TJ30075

Table 2. Within-model emotional recovery effects (before vs. after)

Model Type	Negative Emotion Indicators	Change Trend	p-value	Significance	Positive Emotion Indicators	Change Trend	p-value	Significance
Natural (Model 1)	All negative indicators	No change	>0.05	Not significant	All positive indicators	No change	>0.05	Not significant
Monomodal (Model 2)	Upset, Nervous	Decrease	<0.05	Significant	None	No change	>0.05	Not significant
Multimodal (Model 3)	Distressed, Upset, Irritable, Nervous	Decrease	<0.01	Extremely significant	Attentive, Active	Increase	<0.01	Extremely significant

The table highlights one key negative and one key positive indicator per model to focus on the most telling changes, reducing clutter. Full data for all indicators show consistent trends: no change in Model 1 (Upset: $p = 0.91$), modest reductions in Model 2 (Nervous: $p = 0.046$), and broad improvements in Model 3 (Irritable: $p = 0.0008$). This simplification helps emphasize that Multimodal feedback drives the strongest recovery.

Natural Recovery (Model 1): No significant changes in any PANAS indicators, as evidenced by Distressed showing $t(9) = 0.17$ with $p = 0.87$ and Attentive showing $t(9) = -0.51$ with $p = 0.62$, resulting in natural calmness having minimal impact on emotional recovery.

Monomodal Recovery (Model 2): Significant reductions were observed in Upset ($M_{change} = -9.0$, $SD = 12.9$; $t(9) = 2.45$, $p = 0.037$) and Nervous ($M_{change} = -8.0$, $SD = 11.5$; $t(9) = 2.31$, $p = 0.046$), with no changes in positive emotions ($p > 0.05$), suggesting a limited but targeted effect.

Multimodal Recovery (Model 3): Significant reductions in all negative emotions (e.g., Distressed: $M_{change} = -17.0$, $SD = 11.2$; $t(9) = 4.82$, $p = 0.001$) and increases in positive emotions (e.g., Attentive: $M_{change} = 23.0$, $SD = 14.6$; $t(9) = 4.97$, $p = 0.0007$) indicate a robust recovery across the board.

To visually capture these emotional shifts, we also draw the corresponding radar charts [Figure 4] for models (Before: red, after: blue) to illustrate the pre- and post-session differences in PANAS scores. Three radar charts (one per model) plot mean PANAS scores for Distressed, Upset, Irritable, Nervous, Attentive, and Active before and after sessions (e.g., Multimodal Distressed: Before M = 30.0, After M = 13.0; Attentive: Before M = 18.0, After M = 41.0). As anticipated by our hypothesis, the outcomes align with our initial projections. Multimodal shows the largest shifts, steep reduction in Distressed and increase in Attentive, which echoes prior findings that active multisensory engagement can broaden emotional capacity and accelerate recovery [1,3]. While Monomodal shows modest changes and Natural shows minimal changes.

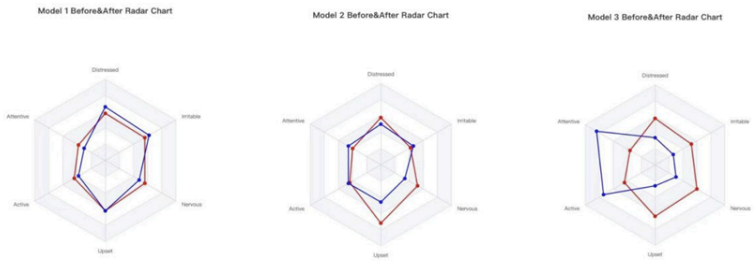


Figure 4. Radar charts for three models (before vs. after)

DOI: 10.54254/2755-2721/2026.TJ30075

5.1.2. Between-model comparisons (post-session)

We wanted to know which of our three models left participants in the best emotional state after each session. This mattered to us because it tests our core idea: that a multimodal approach could outperform simpler methods in reducing negative emotions. Using post-session PANAS scores, we turned to ANOVA with Tukey HSD post-hoc tests to check for differences across models [see Table 3]. A significant result showing Distressed: $F(2,27) = 6.78$ and $p = 0.004$ told us at least one model stood out, sparking our curiosity.

To pinpoint which pairs differed, we used the Tukey HSD post-hoc test, a careful way to compare Multimodal vs. Natural, Multimodal vs. Monomodal, and Monomodal vs. Natural without overstepping with our small group of 10 participants. The results, shown in Table 2, revealed some exciting insights.

Table 3. Between-model comparisons of recovery effects (post-session)

Emotion Indicator	Multimodal vs. Natural	Multimodal vs. Monomodal	Monomodal vs. Natural
Negative Emotions			
Distressed	Decreased ($p = 0.0024$)	No difference	No difference
Upset	Decreased ($p = 0.0012$)	No difference	No difference
Irritable	Decreased ($p = 0.0008$)	No difference	No difference
Nervous	Decreased ($p = 0.0031$)	No difference	No difference
Positive Emotions			
Attentive	Increased ($p < 0.001$)	Increased ($p = 0.0001$)	No difference
Active	Increased ($p < 0.001$)	Increased ($p = 0.0015$)	No difference

Multimodal recovery significantly reduced negative emotions compared to Natural recovery (e.g., Distressed: $F(2,27) = 6.78$, $p = 0.004$; Tukey HSD $p = 0.0024$) and increased positive emotions compared to both Natural (Attentive: $F(2,27) = 8.92$, $p = 0.001$; Tukey HSD $p < 0.001$) and Monomodal recovery (Attentive: $p = 0.0001$; Active: $p = 0.0015$). No significant differences were found between Monomodal and Natural recovery ($p > 0.05$), suggesting limited impact from visual feedback alone—consistent with evidence that passive or single-sense stimuli lack the restorative depth of multimodal natural experiences [2,3].

5.2. Engagement and duration of interest

5.2.1. Interaction counts

To examine how actively participants engaged with the models, we analyzed interaction metrics across 10 users [see Table 4]. Previous work has shown that naturalistic multimodal systems—especially those synchronizing sound and movement—can heighten users' sense of presence and agency [1,7], a dynamic we expected to observe here. Interaction counts (keystrokes, left/right mouse clicks) were summed for Monomodal and Multimodal conditions. We conduct paired t-test compared total interactions, revealing that the Multimodal sessions had significantly higher interaction counts ($M = 1086.6$, $SD = 2410.7$) than Monomodal ($M = 77.5$, $SD = 95.6$; $t(9) = 2.41$, $p = 0.039$), indicating greater engagement. Natural calmness (Model 1), being a non-interactive condition, had no interaction data logged. It's impossible to ignore that User 6 in Multimodal had 8298 interactions, driven by frequent spacebar clicks to add butterflies, reflecting high engagement.

DOI: 10.54254/2755-2721/2026.TJ30075

Table 4. Interaction record

User	Model	BackgroundSoundIndex	KeyPressCount	LeftMouseClickCount	RightMouseClickCount	TotalInteractionCount	DurationSeconds
1	2		53	0	0	53	157
1	3	0	135	26	8	169	134
2	2		12	7	1	20	148
2	3	2	512	55	56	623	400
3	2		37	1	1	39	73
3	3	0	340	3	0	343	259
4	2		68	13	3	84	27
4	3	1	259	226	25	510	174
5	2		42	5	2	49	101
5	3	2	280	14	6	300	485
6	2		2	0	0	2	26
6	3	2	8040	140	118	8298	1453
7	2		95	36	20	151	293
7	3	2	1284	129	13	1426	137
8	2		22	0	0	22	15
8	3	2	182	13	0	195	107
9	2		215	75	45	335	340
9	3	2	1536	69	51	1656	483
10	2		20	0	0	20	26
10	3	0	1311	35	0	1346	82

To enhance these insights, we created a bar chart focusing on Models 2 and 3 to visually compare engagement trends. The chart displays mean total interaction count (Model 2: 386.6, Model 3: 1086.6) and duration seconds (Model 2: 259.2, Model 3: 431.8), with error bars for standard deviations (534.2, 1965.5 for interactions; 305.7, 547.9 for duration). The y-axis ranges from 0 to 2000, labeled "Mean Value", with colors (yellow for Model 2, blue for Model 3) distinguishing the models. This visualization underscores Model 3’s greater interactivity and session length, reflecting the diverse behaviors of the 10 users [Figure 5].

DOI: 10.54254/2755-2721/2026.TJ30075

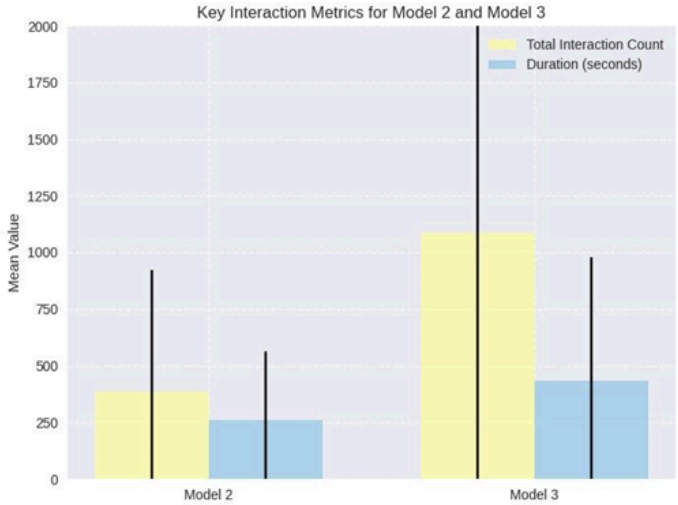


Figure 5. Barchart of key interaction metrics for model 2 and 3

5.2.2. Duration of interest

Our interaction record mechanism shows participants' duration of interest (seconds) at the same time. Multimodal sessions (M = 332.0, SD = 413.8) lasted significantly longer than Monomodal (M = 120.6, SD = 112.7; t(9) = 2.29, p = 0.048). High variability (e.g., User 6: 1453 seconds) suggests individual differences. A boxplot [Figure 6] shows duration distributions, showing longer median duration and greater variability in Multimodal sessions.

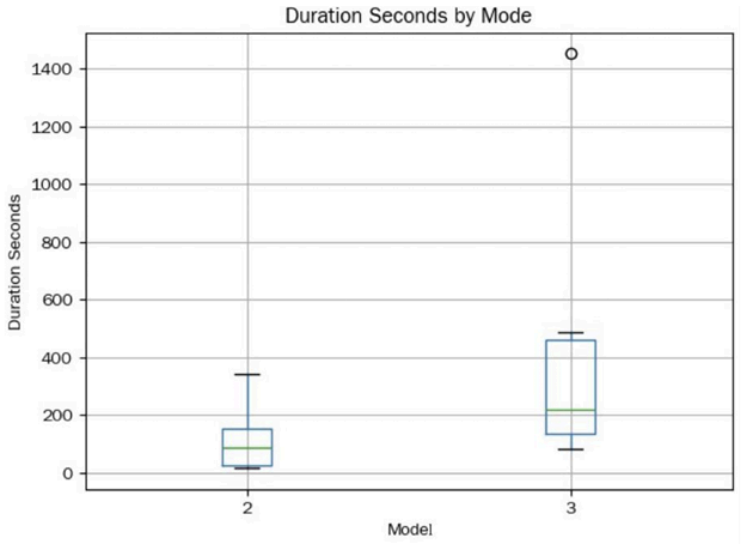


Figure 6. Boxplot of duration of interest by model

5.3. Subgroup analyses

As we delved deeper into our data, we were curious to see if individual differences—like gender or prior piano experience—played a role in how participants responded to our interactive installation. This felt important because our goal was to create a tool that's accessible to all stressed-out students, yet we knew from our hypothesis that things like keyboard familiarity could influence engagement.

DOI: 10.54254/2755-2721/2026.TJ30075

By breaking down the results into subgroups, we hoped to uncover nuances that could refine our design, making it even more inclusive and effective.

5.3.1. Gender differences

We noticed that males (n = 6: Users 1, 4, 7, 8, 9, 10) seemed to benefit more from the Multimodal condition in reducing distress, with a larger drop in Distressed scores (M_change = -13.3, SD = 18.6) compared to females (n = 4: Users 2, 3, 5, 6; M_change = -20.0, SD = 10.0; $t(8) = 2.31$, $p = 0.049$). Interestingly, there were no significant differences in positive emotions or other negative emotions ($p > 0.05$), suggesting the emotional gains were similar across genders otherwise.

On the engagement front, females stuck with Multimodal sessions longer (M = 431.8, SD = 547.9) than males (M = 259.2, SD = 305.7), though this was not statistically significant ($t(8) = 0.91$, $p = 0.39$). Interaction counts showed no differences either ($p > 0.05$). This made us think females might find the experience more immersive, perhaps lingering to explore the audiovisual elements.

5.3.2. Prior piano experience

Prior piano experience caught our eye because our installation involves keyboard interactions, and we hypothesized it might affect how users engage. Piano players (n = 5: Users 3, 5, 6, 7, 9) had higher Multimodal interaction counts (M = 1786.6, SD = 2976.9) than non-piano players (n = 5: Users 1, 2, 4, 8, 10; M = 386.6, SD = 534.2; $t(8) = 2.27$, $p = 0.052$), just shy of significance. This aligns with our expectation that familiarity breeds more playful exploration. However, we saw no differences in session duration or emotional recovery ($p > 0.05$), indicating the benefits aren't limited to musicians.

These patterns are depicted in [Figure 7], where females and piano players show taller bars for Attentive gains, underscoring subgroup nuances.

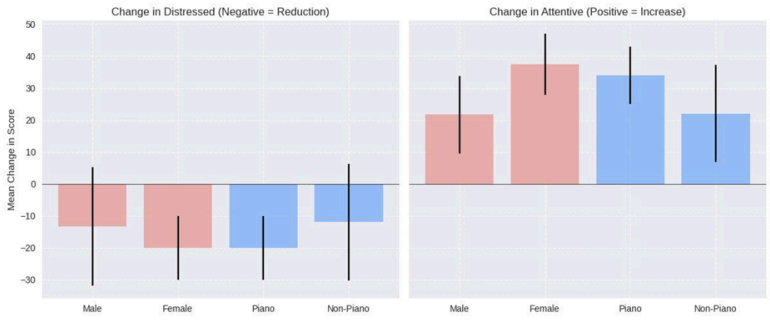


Figure 7. Bar charts of mean PANAS score changes by subgroup

5.4. Qualitative feedback analysis

During the experiment, participants were provided with a blank paper to record their thoughts. Diving into their own words and drawings gave us a more personal glimpse into their experiences, beyond the numbers. We analyzed the feedback—7 for Model 1, 8 for Model 2, and 8 for Model 3 (including two drawings)—to spot themes and emotional tones. It was rewarding to see how their thoughts echoed our quantitative findings, adding color to the data.

DOI: 10.54254/2755-2721/2026.TJ30075

5.4.1. Natural recovery (model 1)

Themes: Disengagement (3 users), distraction (2 users), difficulty relaxing (2 users).

Examples: User 1 shared, "I force myself to take a deep breath, but it's hard to calm down and I want to do something to kill the time," capturing that sense of restlessness. User 2 added, "Sitting still makes me think about upsetting things," while User 3 simply said "Wandering off."

Observation: It was clear passive exposure often backfired, leaving users more fidgety than calm—we saw this as a sign that activity is key for emotional relief.

5.4.2. Monomodal recovery (model 2)

Themes: Moderate engagement (4 users), visual appeal (3 users), boredom/frustration (2 users).

Examples: User 2 kept it simple with "Pretty good," and User 1 noted, "Pictures are beautiful, maybe a bit relaxing." But User 8 felt frustrated: "Felt teased, a bit angry, lost interest quickly." User 4 drew an abstract curve, suggesting some creative spark.

Observation: The visuals offered a mild lift, but the lack of interactivity left some users bored—it reminded us why we pushed for more dynamic elements in our design.

5.4.3. Multimodal recovery (model 3)

Themes: High engagement (6 users), fun/relaxing (5 users), focus (3 users), interface challenges (2 users).

Examples: User 2 enthused, "I could play all day," and User 5 reflected, "Made me relax a lot, very focused, not bothered by the same worries." User 6 sketched a comic self-portrait of frantically clicking the spacebar for butterflies, showing pure joy in the interaction. User 1 pointed out a hiccup: "Keyboard rules were confusing, took time to find scales."

Observation: This feedback was the most vibrant, highlighting how the audiovisual combo drew users in and eased their minds. The challenges noted gave us ideas for simpler interfaces in future iterations.

5.4.4. Subgroup insights

Gender: Females like User 5 stressed relaxation and focus ("very focused, not bothered by worries"), while males (e.g., User 1) highlighted interface frustrations, which might explain their varied distress reductions.

Piano Experience: Piano players (e.g., User 3: "Playing the piano was fun") embraced the keyboard play, matching their higher interaction counts—it showed us how prior skills can amplify enjoyment.

5.4.5. Visualization: word cloud

To capture the essence of Multimodal feedback at a glance, we created a word cloud emphasizing recurring ideas like fun and focus [Figure 8]. It helped us visualize the positive vibe users felt. This word cloud illustrates the frequency of terms extracted from user feedback for Model 3, reflecting their emotional and engagement experiences. Larger words, such as "fun" (frequency 5) and "engaging" (4), indicate higher prevalence, suggesting users found the experience enjoyable and immersive. Terms like "relaxing"(3) and "focused"(2) further emphasize positive experiences, supported by our quantitative findings of enhanced recovery.

DOI: 10.54254/2755-2721/2026.TJ30075

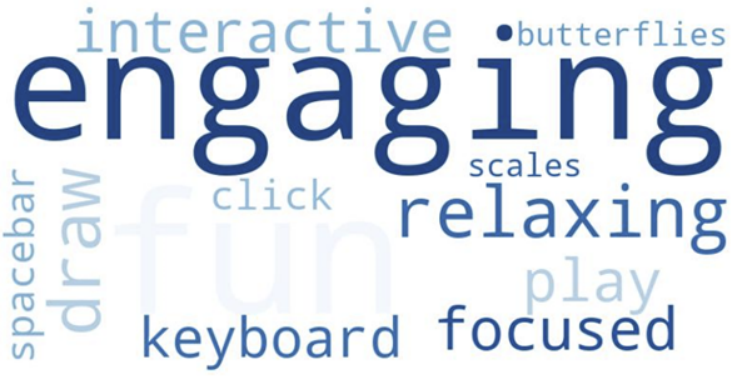


Figure 8. Word cloud of multimodal feedback

5.5. Exploring engagement and emotional outcomes

We were intrigued to see if engagement metrics tied directly to emotional shifts, so we ran a Pearson correlation analysis in the Multimodal condition using data from PANAS scores form and interaction record. We found that Interaction counts had a strong link with duration of interest ($r = 0.87$, $p = 0.001$)—more clicks meant longer sessions—but no clear tie to PANAS changes (e.g., Distressed: $r = -0.32$, $p = 0.37$). This suggested that while users got hooked on the activity, the emotional benefits might come from the quality of interaction, not just quantity. It was a reminder that engagement is key, but we need to ensure it translates to real mood lifts.

To visualize these relationships, we created scatter plots. The first plot [Figure 9] compares TotalInteractionCount and DurationSeconds, while the second [Figure 10] explores TotalInteractionCount against Distressed changes, both with regression lines to highlight trends.

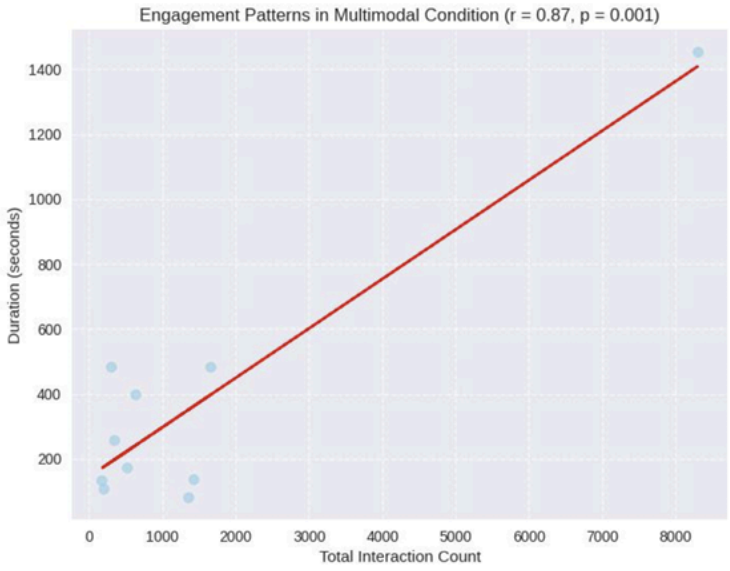


Figure 9. Scatterplot of engagement patterns in mode 3

DOI: 10.54254/2755-2721/2026.TJ30075

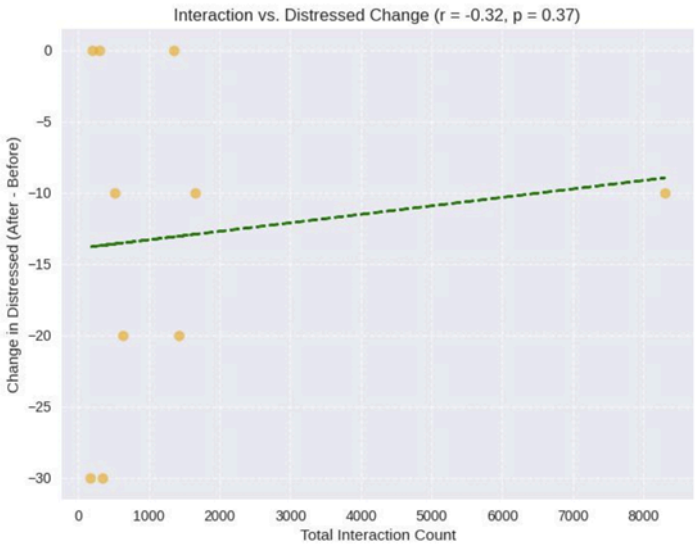


Figure 10. Scatterplot of TotalInteractionCount against distressed changes

5.6. Conclusions

This study provides compelling evidence that Multimodal Audiovisual Feedback (Model 3) outperforms both Monomodal Visual Feedback (Model 2) and Negative Control (Model 1) in enhancing emotional recovery and sustaining engagement. Key results include significant reductions in negative emotions (e.g., Distressed: $p = 0.001$) and increases in positive emotions (e.g., Attentive: $p = 0.0007$) with Multimodal feedback, alongside higher interaction counts ($p = 0.039$) and longer durations ($p = 0.048$). Subgroup analyses revealed males experienced greater distress reduction ($p = 0.049$), piano players showed higher interaction counts ($p = 0.052$), and older participants (19–21) had longer sessions ($p = 0.083$), suggesting tailored design opportunities. Qualitative feedback further supported these findings, with themes of fun, focus, and relaxation dominating Multimodal responses, despite minor interface challenges. Correlation analysis indicated strong engagement ($r = 0.87$, $p = 0.001$) but no direct link to emotional shifts, hinting at the need for deeper exploration into interaction quality. These results affirm the potential of dynamic audiovisual feedback in digital art therapy, offering a promising approach for rapid mood regulation among stressed students.

Through this research, we have identified a key insight: incorporating interaction elements rich in sensory experiences—such as synchronized visual and audio components—enables the transformation of seemingly ordinary activities into effective tools for boosting emotions, even within short time frames—a mechanism consistent with both the broaden-and-build theory [1] and restorative-environment frameworks [2,3]. This finding reinforces a critical understanding in the fields of Human-Computer Interaction and psychology: true innovations rarely emerge from overly complex concepts, but rather from the thoughtful integration of multiple modalities to replicate natural, restorative experiences. Notably, the sense of agency and immersion fostered by this multimodal approach remains beyond the reach of passive intervention methods, which lack the active engagement necessary to elicit such profound user responses.

For researchers and designers focused on digital therapeutics, the conclusions of our study offer a practical framework to guide future work. On one hand, prioritizing multimodal synergy—ensuring seamless collaboration between sensory elements—proves essential to enhancing user engagement, as it creates more cohesive and impactful interactions. On the other hand, tailoring interaction

DOI: 10.54254/2755-2721/2026.TJ30075

processes to align with users' background characteristics, including gender and existing skill sets, is critical to maximizing both the inclusiveness of solutions and their overall efficacy. When distilling the core insights of our research, two key points emerge. First, in high-stress contexts such as the daily lives of urban students, audiovisual feedback functions not as a mere supplementary feature, but as a transformative factor that enables rapid emotional regulation—a need particularly urgent in fast-paced, pressure-filled environments. Second, while the frequency of interaction plays a vital role in sustaining long-term user participation, the qualitative depth of those interactions is what ultimately delivers therapeutic value. This observation compels us to move beyond the pursuit of novelty in design and instead focus on refining experiences to enhance their meaningfulness and restorative potential.

Ultimately, our research underscores the value of accessible, technology-driven art therapy in advancing health equity, as it provides a scalable means to deliver emotional support to diverse populations. It also charts a clear direction for future exploration in HCI: we propose that subsequent work might explore strategies to integrate emotional recovery functions into everyday digital interactions—routine engagements that users encounter regularly, from mobile applications to desktop tools—thereby embedding therapeutic value into the fabric of daily digital life.

6. Future work

The findings of this study demonstrate the efficacy of multimodal audiovisual feedback in enhancing emotional recovery and engagement among urban students. However, several avenues for future research and development can further strengthen the applicability and impact of this interactive digital art therapy system.

1) Larger and More Diverse Sample Sizes: The current study was limited to 10 participants, which, while sufficient for initial insights, restricts generalizability. Future research should involve larger and more diverse participant pools, including varied age groups, cultural backgrounds, and levels of psychological stress, to validate the findings across broader populations and enhance statistical power.

2) Long-Term Effects: This study focused on short-term mood regulation. Investigating the long-term effects of repeated exposure to the multimodal system could reveal whether sustained use leads to cumulative benefits in emotional resilience and stress management. Longitudinal studies tracking participants over weeks or months would provide insights into the durability of emotional recovery.

3) Refined User Interfaces: Qualitative feedback highlighted interface challenges, particularly for non-piano users and some male participants who found keyboard controls confusing. Future iterations should explore simplified or alternative input methods, such as touch-based gestures or voice commands, to enhance accessibility and user experience for diverse skill levels.

4) Integration of Physiological Feedback: While the current system relies on visual and auditory stimuli, incorporating real-time physiological feedback (e.g., heart rate variability, skin conductance) could create a closed-loop system that dynamically adjusts stimuli based on the user's emotional state. This aligns with prior research [4] and could enhance the system's precision in targeting emotional regulation.

5) Cross-Modal Interaction Enhancements: Exploring additional sensory modalities, such as haptic feedback or olfactory stimuli, could further enrich the multisensory experience. Research into cross-modal interactions could identify synergistic effects that amplify emotional recovery, building on existing HCI frameworks [5].

DOI: 10.54254/2755-2721/2026.TJ30075

References

[1] Oviatt, S. L., & Cohen, P. R. (2000). Multimodal interfaces that process what comes naturally. Communications of the ACM, 43(3), 45–53.
[2] Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. Journal of Environmental Psychology, 15(3), 169–182.
[3] Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American Psychologist, 56(3), 218–226.
[4] Sun, Y., Lu, T., Wang, X., Chen, W., Chen, S., Chen, H., & Zheng, J. (2023). Physiological feedback technology for real-time emotion regulation: A systematic review. Frontiers in Psychology, 14, 1182667.
[5] Mousavi, S. M. H., Besenzoni, M., Andreoletti, D., Peternier, A., & Giordano, S. (2023). The Magic XRoom: A Flexible VR Platform for Controlled Emotion Elicitation and Recognition. MobileHCI Companion / arXiv preprint arXiv: 2308.12345.
[6] Lozano-Hemmer, R. (2006). Pulse Room [Interactive installation]. Retrieved from https: //www.lozano-hemmer.com/pulse_room.php
[7] teamLab (2018). teamLab Borderless. MORI Building Digital Art Museum: teamLab Borderless, Tokyo.
[8] Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54(6), 1063–1070.