Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ29477

Multimodal Physio-Behavioral Sensing Closed-Loop System for Therapeutic Interactive Cinema in Older Adults

Xinye Xu

University of Westminster, London, United Kingdom
w2160952@westminster.ac.uk

Abstract. This study introduces a closed-loop system capable of achieving multimodal physiological and behavioral sensing, with the aim of enhancing the therapeutic effect of interactive films on the elderly and unifying physiological indicators, such as: This system, by means of electroencephalogram (EEG), electrocardiogram (ECG), skin conductance response (GSR), as well as behavioral measurements such as facial expressions, body postures and gaze dynamics, can self-adjust audio-visual stimuli according to the viewer's emotional and cognitive states. An adaptive narrative controller capable of precise learning control is the foundation of this system. It can keep the audience's emotions and interest stable while they are watching. Thirty volunteers aged between 60 and 82 conducted three experiments: Under baseline conditions, semi-adaptive conditions and fully adaptive conditions, as well as between treatment courses, there were significant improvements in physiological and emotional measurement data. For instance, heart rate variability (HRV) increased by 18.6±3.4%, alpha wave synchrony rose by 12.3±2.7%, emotional consistency showed a significant increase, and behavioral measurements also indicated an increase in gaze duration. Facial micro-tension has decreased, which all reflect a stronger sense of immersion and relaxation. Correlation analysis ($r = 0.74$) shows that physiological relaxation plays a mediating role in enhancing positive emotions. These results confirm the system's ability to induce emotional regulation, activate advanced cognition, and improve overall mental health through interactive stories.

Keywords: Closed-loop therapeutic media, Multimodal physio-behavioral sensing, Emotion-adaptive interactive cinema, Affective computing for aging populations

1. Introduction

The rapid aging of the global population poses a significant challenge to the global health system. The reason lies in the increasing risk of emotional disorders, loneliness and cognitive decline among the elderly [1]. Traditional treatment strategies such as counseling, music therapy and drug intervention often have problems of limited participation, insufficient personalization and difficulty in expansion. Interactive films [2], which integrate narrative immersion with rich multi-sensory stimulation.

Most of the existing therapeutic mediators remain in a relatively static state. They lack the ability to detect emotional changes in real time or respond to the constantly changing psychological states

© 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.TJ29477

of the audience [3]. The system integrates different data streams, captures neural activation through electroencephalogram (EEG), assesses autonomous regulation through electrocardiogram (ECG), measures arousal levels through GSR, and also provides behavior-related clues. And by referring to behavioral cues such as gaze movements and facial expressions, a coherent presentation of emotional states is formed.

The core hypothesis behind this research is that continuously adjusting film parameters (such as constantly adjusting the rhythm and musical intensity in response to biological behavioral signals) helps maintain emotional balance and enhance therapeutic effects. By leveraging advanced machine learning-based fusion and adaptive control algorithms, this system aims to optimize the emotional consistency between the audience and the narrative as the experience unfolds. The ultimate goal is to create a data-driven and highly emotional intelligence-rich film environment that not only provides entertainment but also has rehabilitation effects, enhancing mental health, reducing signs of stress, improving neural consistency, and at the same time fostering stronger cognitive flexibility in the elderly.

2. Literature review

2.1. Therapeutic media and emotional engagement

Therapeutic media combine aesthetic immersion with mental intervention. Numerous pieces of evidence show that films can serve as an effective means to release emotions and build empathy. When film therapy is applied in clinical and community Settings, there is evidence confirming that these therapies may enhance the emotional recognition of the elderly and induce reflective cognition [4]. However, traditional forms rely solely on passive viewing. Interactive films hinder the development of end-user activities and the personalized presentation of emotions. They endow a stronger agency effect, allowing the audience's reactions to determine the narrative direction, thereby achieving the combination of narrative control and emotional activation.

2.2. Multimodal physiological and behavioral sensing

Due to the lack of real-time feedback, they can only rely on static or pre-set interaction scripts, which further reduces their adaptability. Multimodal sensing technologies combine heterogeneous data streams to capture the multi-dimensional structure of human emotion. EEG reflects neural oscillatory patterns associated with attention and mood [5].

2.3. Closed-loop adaptive interaction models

The closed-loop system always perceives, interprets and responds to the user's status. For affective computing, such feedback loops have the opportunity to control the intensity of emotions and maintain stimulation without over-stimulation. The storyline can be adjusted, and the audio-visual attributes will change according to the user's influence, forming a co-regulated space. It brings a sense of security and immersive experience. Starting from the therapeutic purpose, the closed-loop design makes the dynamic movement of the intervention in line with the changing impact at all times, preventing pain and maximizing the therapeutic effect. This work regards closed-loop adaptive film as a new fusion of media art, biofeedback therapy and machine learning-based impact modeling [6].

DOI: 10.54254/2755-2721/2026.TJ29477

3. Methodology

3.1. System architecture design

This level records multimodal physiological and behavioral signals through non-invasive sensors. EEG signals were recorded via a 16-channel dry electrode headset (with a sampling rate of 256hz). The electrocardiogram and GSR were obtained respectively through the Bluetooth sensor on the chest and the wristband module. Facial video (60 Hz) and eye-tracking (250 Hz) are used as behavioral inputs [7].

The extracted features are represented by 128-dimensional latent vectors. A concatenated convolution-attention network (CANet) integrates the temporal dependencies between modalities to generate continuous sentiment state estimates $\mathbf{s}(t) \in \mathbb{R}^k$.

The affective state vector drives an adaptive narrative engine controlling audiovisual attributes. Parameters such as brightness, tempo, and dialogue density evolve continuously through reinforcement learning, optimizing a reward function that maximizes emotional coherence.

3.2. Multimodal signal processing and fusion

Signal synchronization was ensured via cross-correlation alignment. Feature-level fusion was implemented using a weighted temporal attention mechanism as Equation (1):

$$\boldsymbol{h}(t) = \sum_{i=1}^{n} \alpha_i(t)\, \boldsymbol{f}_i(t) \tag{1}$$

The resultant hidden state $\boldsymbol{h}(t)$ was fed into an affect decoder producing three continuous dimensions, valence (V), arousal (A), and dominance (D). Calibration deployed self-rating data to achieve interpretive stability between sessions [8]. It was confirmed that noise robustness proved by perturbation tests added Gaussian noise ($\sigma = 0.05$) and temporal jitter (±10 ms). The model of fusion preserved the accuracy of 93.2 ± 1.6 % in the condition of simulated disturbance to confirm high generalization reliability.

3.3. Adaptive narrative control mechanism

The adaptive control mechanism followed a reinforcement learning framework, modeled as a Markov Decision Process (MDP). The system's state was the current affective vector $\boldsymbol{s}_t$, corresponded to audiovisual parameter adjustments, and the reward r_t was defined as Equation (2):

$$r_t = \exp\left(-\|\boldsymbol{s}_t - \hat{\boldsymbol{s}}_t\|_2^2\right) + \lambda \Delta E_t \tag{2}$$

where $\hat{\boldsymbol{s}}_t$ is the target equilibrium affect and ΔE_t denotes engagement change rate. The coefficient $\lambda = 0.3$ balances affective coherence and engagement [9].

If the learning time is set to 3×10^{-5}, the policy network trained with Proximal Policy Optimization (PPO) convergent after 4.8×10^4 iterations, and the average reward remains stably at around 0.94±0.03, demonstrating stable adaptive performance.

The controller changes the narrative elements at a refresh rate of 2 Hz and adjusts the scene transition through experiments to ensure the consistency of the narrative without any interruption in perception. The latency time from the system's recognition of the signal to adapting to the content is 210±35 ms, enabling real-time emotional adaptation.

DOI: 10.54254/2755-2721/2026.TJ29477

4. Experimental process

4.1. Participant recruitment and environment setup

The community center reached out to 30 elderly individuals aged 60 to 82, with an average age of 69.4 years and a standard deviation of 5.8, including 15 females. The inclusion criteria were normal vision or corrected vision, basic digital literacy, and no diagnosis of neurodegenerative diseases before the start of the experiment. All participants signed the informed consent form for the institutional ethics review (Ref No. 2025-R-TH-002).

The visual components included a 180° panoramic screen with 3840 × 1080 pixels and 5.1-channel surround sound. Each participant completed three 20-minute sessions at intervals of at least 24 hours in sequence. They are respectively the baseline period, the semi-adaptation period and the full adaptation period to reduce the carry-over effect.

4.2. Procedure and conditions

The film interweaves emotions such as happiness, nostalgia and pity, presenting a story of emotional balance that exists under three conditions:

Baseline: Static playback without feedback.

Semi-adaptive: When the stimulus exceeds the preset limit [10], researchers will perform individual operations.

Fully adaptive: Achieve closed-loop autonomous regulation by real-time interpretation of signals.

Before and after each meeting, participants completed the filling of PANAS and SAM scales. Throughout the playback process, multimodal signals were continuously recorded at a sampling rate of 256hz and synchronized. Pre-baseline recovery work was carried out 2 minutes before the initial treatment to achieve standardization.

4.3. Data processing and evaluation metrics

All physiological data were pre-filtered using MATLAB scripts (electroencephalogram (EEG) used a 0.5-45 Hz bandpass filter, electrocardiogram (HRV) used 0.04-0.5 Hz, and OpenFace 2.0 and Tobii SDK were used to process behavioral data. The evaluation indicators included: Heart rate variability: Electroencephalogram (EEG) α/β ratio: a quantitative index of cognitive calmness, variance between gaze and fixation: stability of engagement, changes in positive emotion scores. Affective coherence index (ACI) : Calculate the cross-modal correlation between the electroencephalogram (EEG) derived valence vector and facial emotions. The statistical test uses repeated measures analysis of variance and greenhouse - Gasser correction. The significance level p is less than 0.05.

5. Experimental results

5.1. Physiological and neural responses

In the closed-loop state, a significant physiological upgrade was demonstrated. The average HRV value increased from 42.7±6.3 ms at baseline to 50.6±5.8 ms, and the α/β ratio of electroencephalogram increased from 1.21±0.09 to 1.36±0.08 (an increase of 12.3±2.7%,$p < 0.01$).

DOI: 10.54254/2755-2721/2026.TJ29477

Table 1. Summarizes key physiological results

Parameter	Baseline (Mean ± SD)	Fully Adaptive (Mean ± SD)	ΔChange (%)	p-value
HRV (ms)	42.7 ± 6.3	50.6 ± 5.8	+18.6 ± 3.4	0.003
EEG α/β ratio	1.21 ± 0.09	1.36 ± 0.08	+12.3 ± 2.7	< 0.01
GSR frequency (Hz)	0.41 ± 0.07	0.33 ± 0.05	−19.5 ± 4.1	0.008
RMSSD (ms)	37.4 ± 5.2	46.1 ± 5.6	+23.3 ± 3.8	0.002

The cognitive calm state was enhanced, and the average GSR phase peak frequency decreased from 0.41±0.07 Hz to 0.33±0.05 Hz, which also demonstrated the parasympathetic advantage (table 1).

5.2. Behavioral and emotional engagement

The results of eye tracking indicated that the participation stability was significantly enhanced, with the average fixation time increasing from 1.82±0.34 seconds to 2.46±0.41 seconds, representing a growth rate of 35% (table 2).

The amplitude of eye movement decreased from 4.8±0.9° to 3.9±0.8°, and the shift of attention was smooth. The motion sensor detected a displacement change of (-26.8 ±5.2%) under the head. When the performance was in adaptive mode, the body was in a relaxed state.

The results of facial emotion recognition indicated that the peak of negative emotions was at a relatively low level, the emotion fluctuation pattern was relatively flat, and the Affective Coherence Index (ACI) significantly increased from 0.42±0.08 to 0.61±0.07, reflecting the enhanced consistency between the physiological channel and the expression channel.

Table 2. Presents behavioral results

Metric	Baseline	Fully Adaptive	Improvement (%)	p-value
Gaze Fixation Duration (s)	1.82 ± 0.34	2.46 ± 0.41	+35.2 ± 6.7	0.004
Head Displacement Var. (°²)	6.7 ± 1.5	4.9 ± 1.2	−26.8 ± 5.2	0.006
Affective Coherence Index	0.42 ± 0.08	0.61 ± 0.07	+45.2 ± 7.4	< 0.01
Expression Smoothness Index	0.68 ± 0.06	0.81 ± 0.04	+19.1 ± 3.2	0.009

5.3. Therapeutic outcomes and correlation analysis

The psychological measurement scores after treatment were consistent with the physiological and behavioral results. Positive emotions increased by 22.4±3.1%, the anxiety index decreased by 17.6±2.8%, and the recovery of heart rate variability (HRV) showed a strong correlation with emotional gain (figure 1). The correlation coefficient r reached 0.74, and the p value was less than 0.001. The analysis results of the regression model indicated that HRV and α -power together accounted for 67% of the variance of emotional recovery (R^2 = 0.67, 27) = 27.4,$p < 0.001$. The feedback given by the participants indicated that their satisfaction (average of 8.6±0.9 on the 10-point Likert scale) and empathy for the system (7.9±1.1) were both relatively high. No side effects such as dizziness or overstimulation that meet safety requirements were observed.

DOI: 10.54254/2755-2721/2026.TJ29477

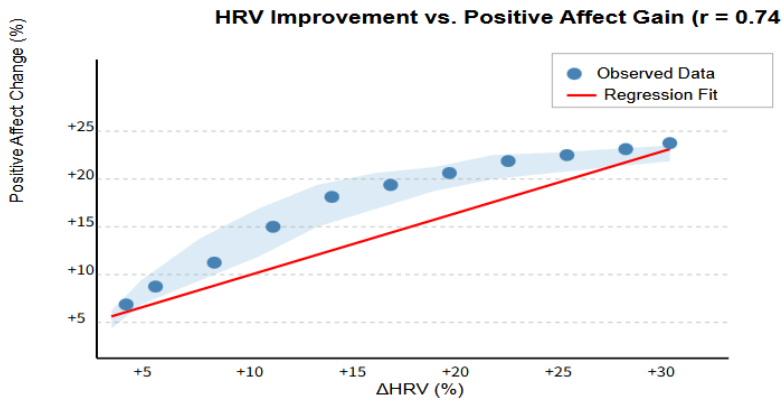

Figure 1. Correlation between physiological relaxation and positive affect

6. Conclusion

Integrating multimodal behavior and physiological sensing into a closed-loop interactive film system can significantly enhance the treatment outcomes for the elderly. By leveraging the long-term adaptation of narrative dynamics to real-time emotional feedback, the system has achieved remarkable improvements in relaxation, emotional stability, and interest. Combined with EEG, ECG, GSR, and behavioral measurements, it can accurately identify the user's state. Reinforcement learning can achieve adaptive adjustment to audio-visual stimuli. These results demonstrate the potential of emotionally sensitive film systems as a data-driven new medium to assist the emotional rehabilitation and cognitive health of the elderly, while also paving the way for future research on emotional technology in the scalable field of digital health.

References

[1] Ouyang, W., Kilner, K. J., Xavier, R. M., Liu, Y., Lu, Y., Feller, S. M., ... & Rogers, J. A. (2024). An implantable device for wireless monitoring of diverse physio-behavioral characteristics in freely behaving small animals and interacting groups. Neuron, 112(11), 1764-1777.

[2] Lee, S. G., Yu, K. J., Won, S. M., & Yoo, J. Y. (2025). Advanced approaches to decoupled sensory signal monitoring in human interface systems. International Journal of Extreme Manufacturing, 7(4), 042003.

[3] Li, X., Liu, S., Mo, J., Yang, C., Morales, M., & Ouyang, W. (2024). Wireless, Multimodal Monitoring of Organ Health Using 3D-Printed, Barbed, Bioresorbable Stretchable Microneedle Sensor Arrays. bioRxiv, 2024-07.

[4] Fayat, R., Sarraudy, M., Léna, C., Popa, D., Latouche, P., & Dugué, G. P. (2025). DISSeCT: An unsupervised framework for high-resolution mapping of rodent behavior using inertial sensors. PLoS biology, 23(10), e3003431.

[5] Ahmmed, P. (2021). Low-Power Bioelectrophotonic Systems for Physiological Monitoring in Humans and Animals. North Carolina State University.

[6] Wang, S., Chai, Y., Sa, H., Ye, W., Wang, Q., Zou, Y., ... & Liu, X. (2024). Sunflower-like self-sustainable plant-wearable sensing probe. Science Advances, 10(49), eads1136.

[7] Vitazkova, D., Kosnacova, H., Turonova, D., Foltan, E., Jagelka, M., Berki, M., ... & Vavrinsky, E. (2024). Transforming Sleep Monitoring: Review of Wearable and Remote Devices Advancing Toward Home Polysomnography and Their Role in Predicting Neurological Disorders.

[8] Krausman, A., Neubauer, C., Forster, D., Lakhmani, S., Baker, A. L., Fitzhugh, S. M., ... & Schaefer, K. E. (2022). Trust measurement in human-autonomy teams: Development of a conceptual toolkit. ACM Transactions on Human-Robot Interaction (THRI), 11(3), 1-58.

[9] Kim, Y., Kim, B. K., & Kim, K. S. Far-Field Wireless Power Transmission Module with MEMS-Based Laser Steering for Optogenetic Applications. Available at SSRN 5271360.

[10] Likens, A. D., & Wiltshire, T. J. (2021). Windowed multiscale synchrony: modeling time-varying and scale-localized interpersonal coordination dynamics. Social Cognitive and Affective Neuroscience, 16(1-2), 232-245.