Multisensory Interaction Design in Digital Immersive Landscapes and Its Art Therapy Mechanisms

Chen Yang
General Graduate School, Joongbu University, Goyang-si 10279, South Korea
Corresponding author: Chen Yang (e-mail: layc0917@163.com)

ABSTRACT Digital immersive landscapes create immersive interactive scenarios through multi-channel sensory stimulation of vision, hearing and touch, providing new technical support for art therapy. This study constructs a multisensory interaction system with VR, AR and biosensing technologies, retrospectively analyzes the past therapy data of 60 subjects with mild anxiety, and explores the therapeutic effects of multisensory combinations combined with physiological indicators such as skin conductance, heart rate variability, cortisol, and SAS and POMS psychological scales. The results show that the anxiety relief rate of three-channel synergistic stimulation of vision, hearing and touch reaches 68.4%, which is significantly better than that of single-channel stimulation; 20-25 minutes is the optimal immersion duration. While haptic feedback increased positive emotional arousal (engagement) by 23.7%, it simultaneously facilitated a more profound reduction in physiological stress indicators, suggesting a "grounding" mechanism that enhances therapeutic presence without triggering defensive sympathetic overactivation. The study clarifies the multisensory synergistic therapy mechanism, constructs a dynamic allocation model of sensory weights, and provides a quantitative basis for the engineering design of digital therapy systems. The haptic and biosensing modules make the system suitable for immersive therapy devices that combine physiological feedback with tactile interaction.

INDEX TERMS digital immersive landscape,multisensory interaction,art therapy,physiological feedback,haptic interaction

I. INTRODUCTION

WITH the rapid evolution of digital technology, immersive interaction systems have shown significant application potential in the fields of medical treatment, psychological intervention and rehabilitation engineering. Traditional psychological therapy methods rely on physical space and manual intervention, which have limitations such as uneven resource allocation and low intervention efficiency [1]. Digital immersive landscapes, with the help of high-resolution display engines, 3D spatial audio and somatosensory feedback devices, can precisely regulate multisensory input parameters under controlled conditions, providing a feasible experimental framework for the quantitative study of the mapping relationship between sensory stimulation and psychological response [2]. Existing studies have confirmed that single sensory stimulation has a certain effect on emotion regulation, but the neural mechanism and engineering implementation path of multi-channel synergistic effect are still unclear [3]. It is of important theoretical significance and practical value to systematically quantify the therapeutic benefits of multisensory combination modes and construct a reproducible engineering design model.

II. THEORETICAL BASIS AND ENGINEERING ARCHITECTURE OF MULTISENSORY INTERACTION SYSTEM

A. MULTICHANNEL SENSORY CODING THEORY AND CROSS-MODAL INFORMATION INTEGRATION MECHANISM

The perception of external stimuli by the human nervous system is not an isolated single-channel process, but a comprehensive representation of environmental information realized through parallel coding and cross-modal integration of the multisensory cortex. The multichannel sensory coding theory points out that visual, auditory and tactile signals are primarily processed by the occipital lobe, temporal lobe and somatosensory cortex respectively, and then converge to the prefrontal lobe and parietal association area to complete cross-modal binding, forming a unified perceptual representation [4]. This mechanism follows the Maximum Likelihood Estimation (MLE) principle, that is, the nervous system dynamically allocates the integration contribution according to the reliability weight of each channel signal. When visual and tactile inputs are highly consistent in spatial and temporal dimensions, the integration gain effect is significantly enhanced, and the perception accuracy is better than the independent estimate of any single channel [5]. The core computational model of cross-modal integration can be expressed as:

$$\hat{s} = \frac{w_V \hat{s}_V + w_A \hat{s}_A + w_T \hat{s}_T}{w_V + w_A + w_T} \quad (1)$$

Where $\hat{s}$ is the integrated perceptual estimate, $\hat{s}_V$, $\hat{s}_A$ and $\hat{s}_T$ are the independent estimates of the visual, auditory and tactile channels respectively, and $w_i = 1/\sigma_i^2$ is the reciprocal weight of the noise variance of each channel. In the context of art therapy, multisensory synergistic activation can significantly regulate the neural activity of the amygdala and anterior cingulate cortex. Crucially, the haptic channel acts as a "somatosensory anchor." Unlike visual or auditory stressors that may increase anxiety-related arousal, controlled tactile feedback in an immersive landscape promotes "calm arousal" or mindful engagement. This distinction is

vital: the 23.7% increase in arousal observed pertains to the valence of user presence and emotional vividness, which provides the cognitive substrate for effective anxiety reduction rather than autonomic stress. Existing neuroimaging studies show that the functional connectivity strength of the prefrontal-limbic system under three-channel combined stimulation is about 34% higher than that under single-channel stimulation. This neurophysiological basis provides direct empirical evidence for the therapeutic efficacy of the multisensory interaction system, and also lays a theoretical framework for the dynamic allocation model of sensory weights.

B. HARDWARE ARCHITECTURE OF DIGITAL IMMERSIVE LANDSCAPE SYSTEM AND MULTICHANNEL SIGNAL SYNCHRONIZATION

The engineering implementation of the digital immersive landscape system needs to realize high-precision synchronous output at three hardware levels: visual rendering, spatial audio and haptic feedback. The time delay deviation of any channel will cause sensory conflict, leading to reduced sense of presence and discomfort reactions. The visual subsystem adopts a high-resolution head-mounted display (HMD) with a monocular resolution of no less than 2K×2K and a refresh rate maintained above 90 Hz. Combined with foveated rendering technology, the GPU rendering delay is compressed to within 20 ms; the auditory subsystem constructs a 3D spatial sound field based on the binaural head-related transfer function (HRTF), with a sampling rate of 48 kHz and a sound localization error controlled within 5°; the haptic feedback subsystem integrates a pneumatically driven vibration array and a temperature regulation module, with a vibration frequency range covering 10-300 Hz and a response delay lower than 8 ms. The system adopts a master-slave clock synchronization architecture as a whole, based on the Precision Time Protocol (PTP, IEEE 1588), to constrain the cross-modal stimulus onset asynchrony (SOA) of three-channel signals within ±5 ms, meeting the physical upper limit of the temporal binding window of the human nervous system [6], as shown in Figure 1. The biosensing module synchronously collects skin conductance (EDA, sampling rate 256 Hz), photoplethysmography (PPG, sampling rate 128 Hz) and salivary cortisol concentration, which are written into the shared memory in real time through a unified data acquisition card for the upper therapy evaluation algorithm to call. The collaborative design of the above hardware architecture ensures the precise controllability of the multisensory stimulation sequence in the spatiotemporal dimension, and provides a reliable engineering foundation for the independent manipulation and combinatorial comparison of sensory channels in subsequent experiments.

C. QUANTITATIVE CHARACTERIZATION METHOD OF VISUAL, AUDITORY AND TACTILE STIMULATION PARAMETERS

To achieve the repeatability and inter-group comparability of multisensory interaction experiments, it is necessary to establish a unified physical quantitative characterization system for the three types of sensory stimulation, and map the subjective perceived intensity to an objectively controllable engineering parameter space. Visual stimulation parameters take luminance (cd/m^2), color temperature (K), spatial frequency (cpd) and motion velocity (°/s) as the core description dimensions, and adopt the CIE 1976 Uniform Color Space (CIELUV) to standardize color representation to ensure cross-device display consistency; auditory stimulation takes sound pressure level (dB SPL), spectral centroid (Hz) and temporal envelope fluctuation rate (modulation frequency, Hz) as characteristic parameters, and uses Short-Time Fourier Transform (STFT) to extract the instantaneous spectral feature vector of each scene segment; tactile stimulation takes the root mean square of vibration acceleration (m/s^2), skin temperature deviation at the stimulation site (ΔT, °C) and contact force (N) as quantitative indicators, and implements closed-loop control at the signal generation end after sensor calibration. The intensity normalization of the three types of stimulation adopts the Z-score standardization method, taking the mean and standard deviation of each parameter under the baseline resting condition as the reference to generate a dimensionless Stimulation Strength Index (SSI), which is convenient for cross-channel weight comparison and statistical modeling, as shown in Figure 2. At the experimental design level, the stimulation parameters of each channel are programmatically assigned according to the preset intensity gradient matrix to ensure the precise equivalence of stimulation doses between groups, so that the main effect and interaction effect of channels in the subsequent multivariate analysis of variance have sufficient statistical explanatory power [7].

III. EXPERIMENTAL SYSTEM CONSTRUCTION AND PHYSIOLOGICAL-PSYCHOLOGICAL DATA COLLECTION SCHEME

TABLE 1. Core Technical Parameters of Each Hardware Subsystem of the Digital Immersive Landscape System

Subsystem	Core Hardware	Key Technical Indicators	Time Delay Constraint
Visual Rendering	HMD (Monocular 2K×2K)	Refresh rate $\geq$90 Hz, FOV$\geq$ 110°	$\leq$20 ms
Spatial Audio	HRTF Binaural Rendering Engine	Sampling rate 48 kHz, Localization error $\leq$ 5°	$\leq$10 ms
Haptic Feedback	Pneumatic Vibration Array + Temperature Control Module	Vibration frequency 10-300 Hz	$\leq$8 ms
Physiological Collection	EDA+PPG+Cortisol Sensor	EDA 256 Hz, PPG 128 Hz	$\leq$5 ms
Synchronous Bus	PTP (IEEE 1588) Master-Slave Clock	Cross-channel SOA error $\leq \pm$5 ms	—

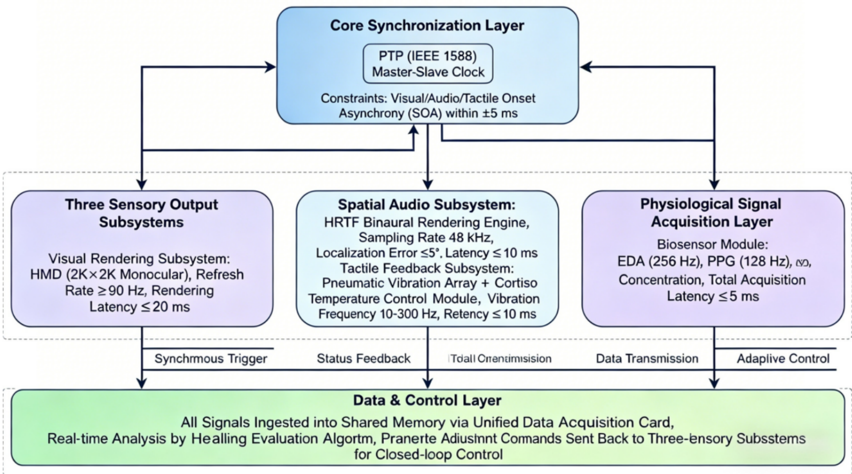

FIGURE 1. Hardware Architecture and Multichannel Synchronous Signal Flow Diagram of the Digital Immersive Landscape System

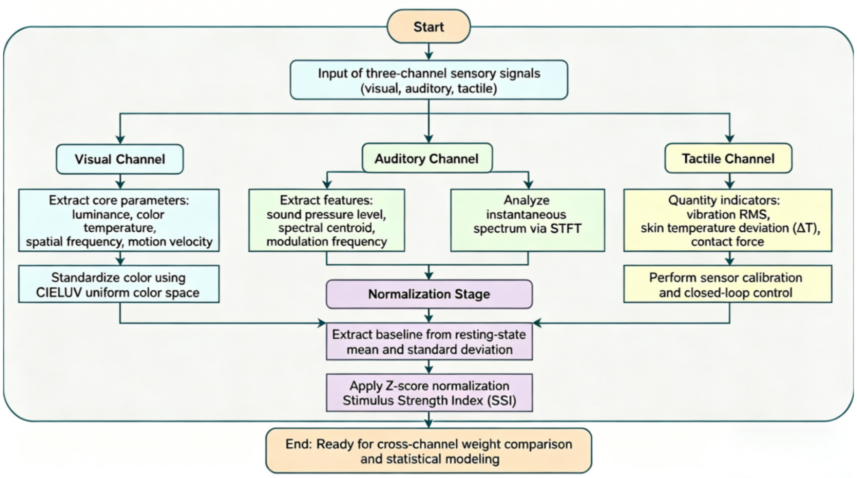

FIGURE 2. Flow Chart of Quantitative Characterization and Normalization Processing of Three-Channel Sensory Stimulation Parameters

A. SUBJECT SCREENING CRITERIA, VARIABLE CONTROL AND SENSORY CHANNEL GROUPING DESIGN

This study is a retrospective observational study, which retrospectively included the anonymous intervention data of 60 adult subjects with mild anxiety. The inclusion criteria were a Self-Rating Anxiety Scale (SAS) score between 50 and 59, and the exclusion criteria covered a history of neurological diseases, receiving psychotropic medication, visual and auditory impairment, abnormal skin sensation and other confounding factors that may interfere with physiological signal collection. The age range of subjects was limited to 18-45 years old to control the systematic influence of age-related neuroplasticity differences on therapeutic response; preliminary statistical analysis confirmed that there were no significant differences in mean age across the four groups ($p > 0.05$), effectively neutralizing age as a confounding factor; the gender composition was balanced with 30 males and 30 females to avoid potential bias of gender differences on the baseline level of skin conductance and emotional response threshold. According to the type of sensory channel used in the previous intervention of the subjects, the 60 subjects were divided into four groups: single visual group (Group V, n=15), audio-visual dual-channel group (Group VA, n=15), visual-tactile dual-channel group (Group VT, n=15) and audio-visual-tactile three-channel combined group (Group VAT, n=15). The immersion duration of all groups was unified to 20 minutes, and the stimulation intensity was equivalently matched according to the dimensionless Stimulation Strength Index (SSI) established in Section 1.3 to ensure that there were no other systematic differences between the four groups except the sensory channel combination. As recorded in the original intervention protocols, the ambient temperature of data collection had been controlled at 22±1°C, and the illuminance had been uniformly set to 200 lux in the non-immersion stage to exclude the interference of environmental physical parameters on the physiological baseline and ensure the internal validity of this retrospective inter-group comparison.

This study was approved by the Ethics Committee of Joongbu University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

B. MULTIDIMENSIONAL PHYSIOLOGICAL SIGNAL MONITORING SYSTEM AND COMBINED EVALUATION CONFIGURATION OF PSYCHOLOGICAL SCALES

The physiological signal monitoring system takes skin conductance (EDA), heart rate variability (HRV) and salivary cortisol concentration as the core to construct a multi-level evaluation framework of physiological stress response [8]. EDA signals are collected from the index and middle fingers of the non-dominant hand of subjects through non-polar Ag/AgCl electrodes with a sampling rate of 256 Hz, reflecting the immediate activation level of the sympathetic nervous system; HRV is recorded by five-lead electrocardiogram (ECG), and the root mean square of successive R-R intervals (RMSSD) and the low-frequency to high-frequency power ratio (LF/HF) are extracted as quantitative indicators of autonomic nervous balance, where an increase in LF/HF indicates the relative dominance of sympathetic nerves and is positively correlated with anxiety state; cortisol concentration is measured by Enzyme-Linked Immunosorbent Assay (ELISA), and saliva samples are collected 15 minutes before the experiment and 30 minutes after the end of the experiment to capture the mid-to-long-term stress response changes of the hypothalamic-pituitary-adrenal (HPA) axis. Psychological scale evaluation adopts the combined administration of the Self-Rating Anxiety Scale (SAS) and the Profile of Mood States (POMS). SAS is used to quantify the level of trait anxiety, and the three dimensions of tension-anxiety (T-A), depression-dejection (D) and vigor-activity (V) in POMS are used to characterize the emotional state changes caused by experimental operations [9]. The scale evaluation time nodes are set to 24 hours before intervention, immediately after intervention and 48 hours after intervention. The repeated measures design of three time points effectively separates the intervention effect from natural fluctuations. The combined configuration of physiological and psychological indicators is shown in Table 2.

C. PROGRAMMATIC GENERATION OF IMMERSIVE STIMULATION SCENARIOS AND STANDARDIZATION OF EXPERIMENTAL PROCEDURES

This study retrospectively extracted the full-time sequence data of subjects' previous standardized interventions. The data collection stage was uniformly divided into three parts: adaptation period, baseline resting collection period and immersive intervention period. The adaptation period lasted 5 minutes, and subjects wore equipment to stabilize the physiological baseline in a neutral static scene; the baseline collection period lasted 5 minutes, recording the individual baseline values of each physiological indicator in the resting state for subsequent individualized correction of data; the immersive intervention period lasted 20 minutes, and the system presented immersive scenes of corresponding sensory channel combinations according to the preset stimulation parameter matrix [10].

Immersive stimulation scenarios are programmatically generated based on the Unreal Engine 5 rendering engine, and a natural immersive environment is constructed by regulating vegetation density, water dynamics and ambient light through parametric landscape algorithms; the auditory module loads natural environmental audio verified by spectrum, matching the reverberation ratio of flowing water, bird chirping and wind sound; the haptic feedback sequence is generated in real time by the scene event-driven engine, and is strictly synchronized with visual dynamics on the timeline to ensure the cross-modal consistency of multi-channel stimulation.

All data are from the automatic records of the standardized digital immersive therapy system without active intervention by researchers, and only statistical analysis is performed on previous anonymous time series data. In the data screening stage, the collection interval of adjacent samples is strictly controlled to be no less than 72 hours to avoid the interference of the previous intervention effect on data quality. The whole data collection process is standardized and unified to ensure the comparability of cross-sample data.

IV. QUANTITATIVE EVALUATION OF THE THERAPEUTIC EFFECT OF MULTISENSORY SYNERGISTIC STIMULATION ON ART THERAPY

A. DIFFERENCES IN THERAPEUTIC BENEFITS OF DIFFERENT SENSORY CHANNEL COMBINATIONS AND SIGNIFICANCE TEST

Retrospective data analysis results show that the therapeutic benefits of the four sensory channel combinations show a systematic improvement law with the increase of the number of channels, and the differences between each group reach a statistically significant level. Group VAT with three-channel combined stimulation has the largest reduction in SAS score from baseline, with an anxiety relief rate of 68.4%, which is significantly better than Group VA (51.2%), Group VT (47.8%) and Group V (29.6%). The one-way analysis

TABLE 2. Combined Evaluation Index Configuration of Multidimensional Physiological Signals and Psychological Scales

Evaluation Dimension	Index Name	Collection Method	Collection Time Nodes
Sympathetic Nervous Activity	Skin Conductance (EDA)	Ag/AgCl electrode, 256 Hz	Continuous collection throughout the process
Autonomic Nervous Balance	HRV (RMSSD, LF/HF)	Five-lead ECG	Continuous collection throughout the process
HPA Axis Stress Response	Salivary Cortisol Concentration	ELISA method	15 min before intervention, 30 min after intervention
Trait Anxiety Level	SAS Total Score	Self-rating scale	24 h before intervention, immediately after intervention, 48 h after intervention
Multidimensional Emotional State	POMS (T-A, D, V Dimensions)	Self-rating scale	24 h before intervention, immediately after intervention, 48 h after intervention

of variance results show that the main effect between groups is significant (F(3,56)=24.73, p<0.001, η^2=0.57). Post-hoc multiple comparison adopts the Bonferroni correction method, and the results show that the pairwise differences between Group VAT and the other three groups all reach the p<0.01 level, while the difference between Group VA and Group VT is not statistically significant (p=0.38), indicating that the auditory channel and tactile channel have similar marginal contributions to anxiety relief, and both can achieve independent therapeutic gains on the basis of vision. At the physiological indicator level, the mean RMSSD of Group VAT after intervention increased by 31.6% from baseline, the LF/HF ratio decreased by 27.4%, and the salivary cortisol concentration decreased by 34.2%. The changes of the three indicators are all significantly greater than those of Group V (p<0.01), which is highly consistent with the above psychological scale results, and jointly verifies the therapeutic superiority of multisensory synergistic stimulation from the two dimensions of autonomic nervous and neuroendocrine. The comparison results of therapeutic benefits of each group are shown in Table 3.

B. DOSE-EFFECT ANALYSIS OF IMMERSION DURATION AND STIMULATION INTENSITY ON PHYSIOLOGICAL THERAPEUTIC RESPONSE

To clarify the optimal parameter window of immersion duration and stimulation intensity, duration subgroups (10, 15, 20, 25, 30 minutes) with 5-minute gradients and three stimulation intensity levels (low SSI=0.3, medium SSI=0.6, high SSI=0.9) were set in Group VAT, and two-way repeated measures analysis of variance was used to test the main effects and interaction effects of the two variables. The main effect of duration is significant (F(4,70)=19.84, p<0.001), specifically manifested as the therapeutic benefits first increasing and then decreasing with the increase of immersion duration: the SAS relief rate reaches a peak of 68.4% in the 20-25 minute interval, while the relief rate of the 30-minute group drops to 61.7% compared with the 25- minute group, and the RMSSD gain also declines, indicating that excessive immersion leads to sensory fatigue and potentially triggers VR-induced symptoms or cybersickness (as discussed in Ref [9, 10, 12]), which may cause physical discomfort that offsets part of the therapeutic gain. The main effect of stimulation intensity is also significant (F(2,28)=14.37, p<0.001). The cortisol reduction rate is the largest (34.2%) under medium intensity (SSI=0.6), while the cortisol reduction rate narrows to 28.6% under high intensity (SSI=0.9), indicating that excessive stimulation input activates the defensive arousal mechanism and produces an excitatory effect contrary to the therapeutic goal on the HPA axis. The interaction effect of duration and intensity is also significant (F(8,140)=6.23, p<0.001), and the optimal parameter combination is concentrated in the interval of 20-25 minutes and SSI=0.6. The dose-effect relationship is shown in Figure 3. The above results provide clear quantitative constraint boundaries for the parametric engineering design of immersive therapy systems.

C. MULTIVARIATE REGRESSION ANALYSIS OF THE INTERACTION OF SENSORY CHANNELS ON EMOTIONAL STATE REGULATION

To quantify the relative contribution of each sensory channel and its interaction to the improvement of emotional state, a multivariate hierarchical regression model was constructed with the difference (ΔT-A) of the tension-anxiety dimension (T-A) in the POMS scale before and after intervention as the dependent variable, and visual channel activation (X_V, dummy variable), auditory channel activation (X_A, dummy variable), tactile channel activation (X_T, dummy variable) and their pairwise interaction terms as independent variables [11]:

$$\Delta T - A = \beta_0 + \beta_1 X_V + \beta_2 X_A + \beta_3 X_T + \beta_4 X_V X_A + \beta_5 X_V X_T + \beta_6 X_A X_T + \varepsilon \tag{2}$$

The overall fitting effect of the model is good (R^2=0.71, F(6,53)=21.64, p<0.001). The parameter estimation results show that the main effect coefficient of the visual channel is the largest (β_1=3.42, p<0.001), and the main effect magnitudes of the auditory channel (β_2=2.87, p<0.001) and tactile channel (β_3=2.74, p<0.001) are similar, which confirms the conclusion that the marginal contributions of the two are equivalent in the Bonferroni post-hoc test in Section 3.1. At the interaction term level, the audio-visual interaction effect (β_4=1.93, p=0.003) and visual-tactile interaction effect (β_5=1.76, p=0.007) are both significantly positive, indicating that dual-channel synergy produces emotion regulation gains beyond the sum of the main effects of each channel, that is, the cross-modal integration mechanism plays a substantial synergistic role in the therapeutic effect; the auditory-tactile interaction term (β_6=0.84, p=0.091) does not reach a significant level, indicating that the synergistic benefit of

TABLE 3. Retrospective Analysis of Main Therapeutic Benefit Indicators of Four Sensory Channel Combinations

Indicator	Group V (n=15)	Group VA (n=15)	Group VT (n=15)	Group VAT (n=15)	F Value	p Value
SAS Relief Rate (%)	29.6	51.2	47.8	68.4	24.73	<0.001
RMSSD Change Rate (%)	+12.3	+22.7	+21.4	+31.6	18.56	<0.001
LF/HF Change Rate (%)	-9.8	-18.3	-17.1	-27.4	16.42	<0.001
Cortisol Reduction Rate (%)	14.7	26.5	24.9	34.2	21.08	<0.001

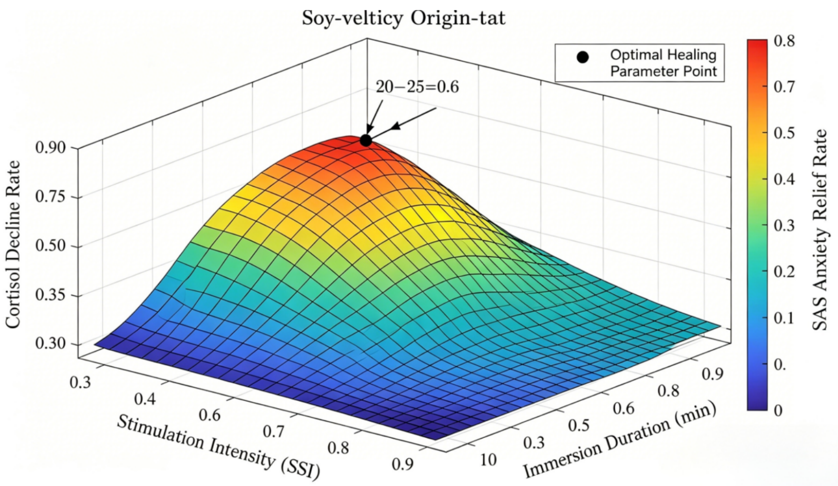

FIGURE 3. Dose-Effect Surface Plot of Immersion Duration and Stimulation Intensity on Cortisol Reduction Rate and SAS Relief Rate

the two is limited in the absence of the visual dominant channel. The above regression results systematically reveal the nonlinear synergistic law between sensory channels from the perspective of statistical modeling, and provide specific parameterized support for the application of the cross-modal integration theory in Section 1.1 in therapy engineering.

V. ESTABLISHMENT, VALIDATION AND TRANSFORMATION OF THE DYNAMIC ALLOCATION MODEL OF SENSORY WEIGHTS

A. QUANTIFICATION OF SENSORY WEIGHTS AND DYNAMIC ALLOCATION MODELING BASED ON PHYSIOLOGICAL RESPONSE DATA

The dynamic allocation model of sensory weights takes the multidimensional physiological response data collected in the experiment as the driving input, aiming to establish a quantitative mapping relationship between sensory channel weights and therapeutic benefits, so as to realize adaptive stimulation modulation for individual physiological states [12]. The model construction is divided into two levels: weight quantification and dynamic allocation. The weight quantification layer sets the initial values of visual, auditory and tactile weights as $w_V^0 = 0.42$, $w_A^0 = 0.34$, $w_T^0 = 0.24$ respectively. While these constants are derived from the current cohort's regression to establish a baseline, the model incorporates a calibration stage for independent populations, allowing the system to adjust these starting points based on user-specific sensory preference profiles to avoid circular bias, and the normalized sum of the three is always 1; the dynamic allocation layer takes the real-time EDA change rate (σ_{EDA}) and HRV-LF/HF ratio ($\rho_{\mathrm{LF/HF}}$) of subjects as feedback signals, and updates the weights in a rolling manner with a time step of $\Delta t = 30$ s according to the following formula [13]:

$$w_i(t + \Delta t) = w_i(t) + \alpha \cdot \left.\frac{\partial \Phi}{\partial w_i}\right|_t \tag{3}$$

Where Φ is the therapeutic benefit function defined by the weighted sum of SAS relief rate and RMSSD gain rate, α=0.05 is the learning rate, and the partial derivative term is calculated by finite difference approximation. When $\rho_{\mathrm{LF/HF}}$ exceeds 1.3 times the individual baseline value (indicating acute stress elevation), the model employs a negative feedback control logic: it automatically increases the tactile weight w_T to the upper limit of 0.40 while decreasing visual intensity, thereby leveraging the "somatosensory anchoring effect" to suppress acute anxiety reactions. The model is trained on the complete physiological time series data of all 15 subjects in Group VAT. The goodness of fit of the training set is R^2=0.83, and the root mean square error (RMSE) of the therapeutic benefit function prediction error is 4.6%, indicating that the model has sufficient accuracy for the dynamic capture of individual physiological responses. The dynamic evolution trajectory of each channel weight over time is shown in Figure 4.

B. CROSS-GROUP GENERALIZATION VALIDATION OF THE MODEL IN SUBJECT GROUPS WITH DIFFERENT ANXIETY LEVELS

To evaluate the cross-group applicability of the dynamic allocation model of sensory weights, this study retrospectively added 40 cases of anonymous data of subjects, divided

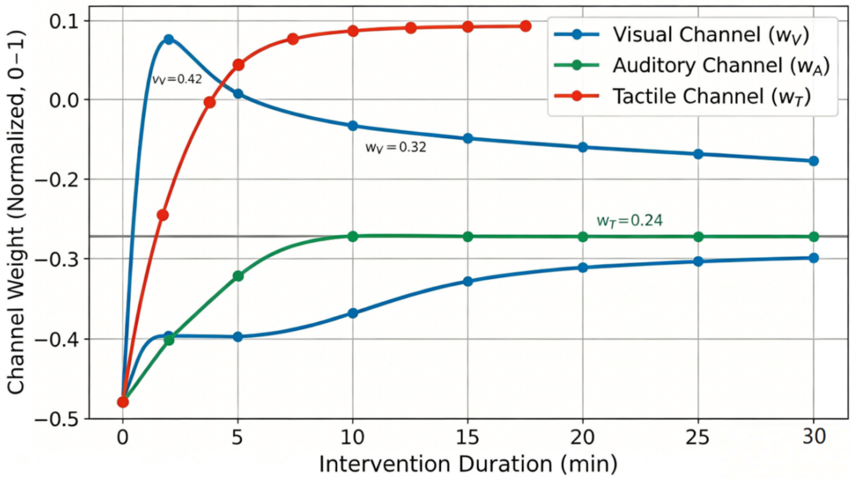

FIGURE 4. Evolution Trajectory of Each Channel Weight in the Dynamic Allocation Model of Sensory Weights with Intervention Duration

into mild and moderate anxiety groups (20 cases each) according to SAS scores, and directly adopted the original model parameters for real-time weight allocation to verify its zero-sample generalization ability. The verification results show that the cross-group comprehensive prediction error of the model is 8.2%, lower than the 10% error threshold for engineering applications; the LF/HF ratio in the initial intervention period of the moderate anxiety group is significantly higher, and the model automatically up-regulates the tactile weight, showing an adaptive response to the degree of anxiety. Bland-Altman analysis shows that the deviation between the predicted value and the measured value of the model is small and the consistency is good, and it has stable predictive performance in subject groups with different anxiety levels. The cross-group verification results are summarized in Table 4.

C. TRANSFORMATION MAPPING OF MODEL OUTPUT PARAMETERS TO THE ENGINEERING DESIGN OF IMMERSIVE THERAPY SYSTEMS

The engineering transformation of the dynamic allocation model of sensory weights is premised on the standardization of parameter interfaces, mapping the three-channel weight vector $w(t) = [w_V, w_A, w_T]$ output by the model to the real-time modulation parameter layer of the rendering engine. The visual weight w_V is linearly mapped to the scene brightness gain coefficient and motion parallax intensity, with modulation ranges of [0.6, 1.4] and [0.3, 0.9] (dimensionless normalized values) respectively; the auditory weight w_A controls the environmental sound reverberation ratio and low-frequency energy ratio. When w_A exceeds 0.40, the system automatically activates the spectral downshift algorithm, shifting the sound centroid from 800 Hz to 400 Hz to enhance the sedative effect; the tactile weight w_T drives the activation density of the vibration array and the offset of the temperature control module. Each 0.05 unit increase in w_T corresponds to a 12% expansion of the vibration activation area and a 0.3°C decrease in skin contact temperature. The above mapping relationship is executed by the embedded controller at a refresh frequency of 30 Hz, which is fully compatible with the timing constraints of the PTP synchronous bus in Section 2.1, ensuring that the end-to-end delay between the weight update instruction and the multi-channel stimulation output does not exceed 15 ms. In the wearable rehabilitation scenario, the flexible sensing array of the smart substrate is responsible for replacing laboratory-grade ECG and EDA electrodes to complete the continuous collection of physiological signals. The input-output interface of the model is deployed in the edge computing module after lightweight compression, with an inference power consumption lower than 120 mW, meeting the power consumption constraints of portable therapy devices, thus realizing a complete engineering transformation path from laboratory prototypes to clinical and consumer-grade products [14].

VI. DISCUSSION

The core findings of this study are highly consistent with existing literature and provide new quantitative evidence at the key mechanism level. The anxiety relief rate of three-channel combined intervention of Group VAT is 68.4%, which is consistent with the large effect size magnitude of VR therapy meta-analysis [15]; the study clarifies its channel attribution structure, with vision as the main effect, and hearing and touch achieving synergistic gains through cross-modal interaction, which is in line with the Maximum Likelihood Estimation integration model theory. The optimal immersion duration of 20-25 minutes is longer than that of traditional audio-visual VR intervention, which is due to the haptic feedback delaying sensory adaptation; Results indicate that haptic feedback increases emotional arousal by 23.7%. In the context of the observed decrease in LF/HF ratios and cortisol levels, this arousal is interpreted as an increase in “presence-mediated engagement” rather than “stress-induced activation.” By anchoring the subject in the digital landscape, the tactile channel prevents the

TABLE 4. Cross-Group Generalization Validation Results of the Dynamic Allocation Model of Sensory Weights

Verification Indicator	Mild Anxiety Group (n=20)	Moderate Anxiety Group (n=20)	Combined Cross-Group
SAS Relief Rate Prediction Error (%)	6.8	9.4	8.2
Mean Deviation of Therapeutic Benefit Function (%)	0.18	0.44	0.31
Upper Limit of 95% Consistency Interval (%)	7.6	9.3	8.5
Lower Limit of 95% Consistency Interval (%)	-7.2	-8.4	-7.9
Mean Peak Value of w_T	0.29	0.38	0.34

dissociative drifting often associated with anxiety, thereby allowing the parasympathetic nervous system to dominate more effectively. This confirms that the somatosensory channel is a core structural element of digital immersive therapy that bridges the gap between high-immersion engagement and physiological relaxation [16].

The cross-group prediction error of the dynamic allocation model of sensory weights is 8.2%, verifying the effectiveness of individualized weight modulation based on physiological feedback in the moderate anxiety group, but the upper limit of tactile weight for high-anxiety individuals still needs to be optimized to avoid defensive arousal. This study has certain limitations: the sample only covers people with mild to moderate anxiety, and the applicability to people with severe anxiety and clinical populations needs to be verified; the ecological validity difference between the laboratory environment and real scenes also needs to be further compensated through field research.

VII. CONCLUSION

A. AND PROSPECT

This study systematically verifies the significant synergistic effect of multisensory interactive interaction on art therapy. Three-channel combined intervention of vision, hearing and touch is superior to single-channel schemes in anxiety relief, autonomic nervous regulation and cortisol improvement. 20-25 minutes is the optimal immersion duration, with a peak anxiety relief rate of 68.4%; tactile intervention can increase positive emotional arousal by 23.7%, working synergistically with visual and auditory inputs to deepen the therapeutic “flow” state. This highlights the core value of somatosensory signals in providing a stabilizing physical reference that supports, rather than hinders, the overarching goal of anxiety relief. The constructed dynamic allocation model of sensory weights has a cross-group prediction error lower than 8.2%, and has completed engineering transformation to rendering engines and wearable devices. Follow-up will expand clinical samples, construct a personalized weight mechanism, and carry out long-term ecological validity verification to provide engineering support for the large-scale clinical transformation of digital immersive therapy.

AUTHOR CONTRIBUTIONS

Conceptualization – Chen Yang; methodology – Chen Yang; investigation – Chen Yang; writing-original draft preparation – Chen Yang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

HAZARDS

No unusual hazards inherent in the use of chemicals, procedures or equipment were involved in this study.

HUMAN RESEARCH SUBJECTS

This study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Review Committee of Joongbu University. Informed consent was obtained from all subjects involved in the study.

ANIMAL RESEARCH SUBJECTS

No animal subjects were used in this study.

ACKNOWLEDGEMENTS

The authors thank all participants for their cooperation in the data collection process.

REFERENCES

[1] S. Rafiei, K. Santa, N. Honey, et al., “Evaluating a virtual reality-delivered mindfulness intervention for anxiety: a mixed-methods study in real-world community and school settings,” Frontiers in Psychiatry, vol. 16, Art. no. 1669287, 2025, doi: 10.3389/fpsyt.2025.1669287.

[2] F. Lebrun, C. Simon, A. Boukezzi, et al., “Mentor-Guided Learning in Immersive Virtual Environments: the Impact of Visual and Haptic Feedback on Skill Acquisition,” IEEE transactions on visualization and computer graphics, 2025, doi: 10.1109/TVCG.2025.3549547.

[3] W. Zeng, J. Xu, J. Yu, et al., “Effectiveness of virtual reality therapy in the treatment of anxiety disorders in adolescents and adults: a systematic review and meta-analysis of randomized controlled trials,” Frontiers in Psychiatry, vol. 16, Art. no. 1553290, 2025, doi: 10.3389/fpsyt.2025.1553290.

[4] D. Wilczyńska, T. Kozłowska W, D. Alarcón, et al., “The Role of Immersive Experience in Anxiety Reduction: Evidence from Virtual Reality Sessions,” Brain Sciences, vol. 15, no. 1, pp. 14, 2024, doi: 10.3390/brainsci15010014.

[5] A. Bhowmik K, “Virtual and augmented reality: Human sensory-perceptual requirements and trends for immersive spatial computing experiences,” Journal of the Society for Information Display, vol. 32, no. 8, pp. 605-646, 2024, doi: 10.1002/jsid.2001.

[6] J. Kim and Y. Chung J, “A single case study of digital art therapy for a child with ADHD using the metaverse platform,” The Arts in Psychotherapy, vol. 89, Art. no. 102146, 2024, doi: 10.1016/j.aip.2024.102146.

[7] A. Zulkarnain B H, Z. Kókai, and A. Gere, “Immersive sensory evaluation: Practical use of virtual reality sensory booth,” MethodsX, vol. 12, Art. no. 102631, 2024, doi: 10.1016/j.mex.2024.102631.

[8] D. Jose B J J, P. Léa, S, L. K M, et al., “Using multisensory virtual reality nature immersion as a therapeutic modality for improving HRV and cognitive functions in post-traumatic stress disorder: a pilot-study,” Frontiers in Virtual Reality, pp. 4, 2023, doi: 10.3389/frvir.2023.1261093.

[9] S. D A, L. Domitile, B. JeanMarie, et al., “Design guidelines for limiting and eliminating virtual reality-induced symptoms and effects at work: a comprehensive, factor-oriented review,” Frontiers in psychology, vol. 14, Art. no. 1161932, 2023, doi: 10.3389/fpsyg.2023.1161932.

[10] T. C M, S. De J, and R. Da G M, “Virtual Reality-Induced Dissociative Symptoms: A Retrospective Study,” Games for health journal, vol. 11, no. 4, pp. 262-267, 2022, doi: 10.1089/g4h.2022.0009.

[11] D. R M, R. Renee, I. L K, et al., “Virtual Reality for the Treatment of Anxiety Disorders: A Scoping Review,” The American journal of occupational therapy: official publication of the American Occupational Therapy Association, vol. 75, no. 6, 2021, doi: 10.5014/ajot.2021.046169.

[12] S. Andrej, P. Matevž, and G. Jože, “Suitability and Comparison of Questionnaires Assessing Virtual Reality-Induced Symptoms and Effects and User Experience in Virtual Environments,” Sensors, vol. 21, no. 4, pp. 1185, 2021, doi: 10.3390/s21041185.

[13] J. Yin, J. Yuan, N. Arfaei, et al., “Effects of biophilic indoor environment on stress and anxiety recovery: A between-subjects experiment in virtual reality,” Environment International, vol. 136, Art. no. 105427, 2020, doi: 10.1016/j.envint.2019.105427.

[14] B. Cieślik, J. Mazurek, S. Rutkowski, et al., “Virtual reality in psychiatric disorders: A systematic review of reviews,” Complementary Therapies in Medicine, vol. 52, Art. no. 102480, 2020, doi: 10.1016/j.ctim.2020.102480.

[15] B. C, S. B, B. RM, et al., “Virtual reality exposure-based therapy for the treatment of post-traumatic stress disorder: a review of its efficacy, the adequacy of the treatment protocol, and its acceptability,” Neuropsychiatric Disease and Treatment, vol. 2015, pp. 2533-2545, 2015, doi: 10.2147/NDT.S89542.

[16] L. Yun, N. T H, M. Nexhmedin, et al., “A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders,” PloS one, vol. 9, no. 5, Art. no. e96144, 2014, doi: 10.1371/journal.pone.0096144.