Developing a Dynamic Music and Dance Coordination Multi-Agent System Based on Real-Time Audience Feedback Analysis

Yue Yu, Xi Zhang
Art Department of Qinhuangdao Vocational and Technical College, Qinhuangdao 066001, Hebei, China
Corresponding author: Xi Zhang (e-mail: 16603351381@163.com)

ABSTRACT Within the context of deep integration between digital entertainment and intelligent interaction, traditional music and dance systems fail to meet contemporary demands for immersive artistic experiences. Real-time audience feedback contains the emotional preferences, aesthetic tendencies, and willingness to participate of the audience, and is the core data resource driving the dynamic optimization of performance content. Multi-agent systems, relying on the advantages of distributed collaboration, can achieve flexible adaptation and real-time response to music generation and dance choreography. This paper develops a dynamic music and dance coordination multi-agent system based on real-time audience feedback analysis, aiming to improve the adaptive collaboration between audience emotion recognition, music generation, dance choreography, and performance control.

INDEX TERMS real-time audience feedback,multi-agent system,music-dance coordination,dynamic generation,interactive performance

I. INTRODUCTION

A. RESEARCH BACKGROUND

MUSIC and dance, as core art forms for human emotional expression and cultural dissemination, are undergoing profound changes in their performance patterns, particularly as intelligent technology enables the rhythmic simulation of fiber morphology and the complex spatial distribution inherent in the industry. This integration allows for the digital preservation of traditional craftsmanship by mapping the repetitive, percussive cycles of antique loom operations—specifically the 'beat-up' and 'shuttle-flying' motions—onto the rhythmic structures of dance choreography. The Overall Layout Plan for the Construction of Digital China clearly proposes to promote the deep integration of digital technology and cultural arts, and cultivate new cultural formats such as digital art and intelligent performance. While intelligent music and dance performance systems are increasingly applied to virtual idols and interactive exhibitions, they still face structural challenges in accurately simulating the complex fiber morphology and rhythmic mechanical movements characteristic of the e industry. These limitations hinder the seamless multimodal integration required to digitally preserve and artistically express the intricate spatial distribution of traditional craftsmanship[1-3].

B. RESEARCH SIGNIFICANCE

1) Theoretical Significance

(1) Build a theoretical framework for real-time audience feedback driven collaborative generation of music and dance, clarify the mapping relationship between feedback information and specific performance variables (e.g., tempo, syncopation, and spatial trajectory), and enrich the theoretical system of intelligent art creation;

(2) Propose a cross modal collaboration mechanism based on multi-agent systems, improve the application theory of multi-agent systems in the field of artistic creation, and provide methodological support for collaborative generation of multimodal content;

(3) Establish a correlation model between audience feedback emotion intention recognition and dynamic optimization of artistic expression, clarify the intrinsic mechanism of real-time interaction in enhancing artistic experience, and expand the theoretical boundary of intelligent interaction and artistic communication[4].

2) Practical Significance

(1) Develop a multi-agent system that integrates real-time feedback collection, intelligent analysis, and collaborative generation, providing technical support for fields such as digital entertainment, intelligent performance, and interactive art, and solving problems such as insufficient interactivity and collaboration in traditional systems;

(2) Drive dynamic optimization of performance content through audience feedback, enhance the personalization, immersion, and sense of participation of artistic performances, and meet the contemporary audience's demand for high-quality artistic experiences;

(3) Provide intelligent creative assistance tools for artists, lower the technical threshold for cross modal art creation, stimulate artistic innovation vitality, and promote the diversified development of digital art formats;

(4) Expand the application scenarios of multi-agent systems and real-time feedback analysis technology, providing reference for technological innovation in fields such as intelligent interaction, virtual simulation, and cultural and technological integration[5-6].

C. CURRENT RESEARCH STATUS AT HOME AND ABROAD

1) Current Research Status Abroad

Research in real-time feedback analysis, multi-agent collaboration, music and dance generation, and other fields started earlier abroad, mainly focusing on three directions:

(1) Real time audience feedback analysis research has developed an emotion recognition system based on physiological signals (such as heart rate, skin conductance) and behavioral data (such as facial expressions, body movements) to achieve real-time capture of audience emotional states;

(2) Research on multi-agent collaboration technology has applied multi-agent systems to scenarios such as robot dance choreography and virtual character interaction, achieving simple action collaboration and behavioral response;

(3) Research on music and dance generation, based on deep learning, reinforcement learning and other technologies, has developed an automatic music generation and dance action choreography system that can generate artistic content that conforms to specific styles.

The advantages of foreign research lie in its profound technical foundation and rich application scenarios, but there are also shortcomings:

Most systems lack comprehensive analysis of multimodal audience feedback and rely solely on feedback information;

Multi intelligent body collaboration is mostly limited to a single mode (such as only dance action collaboration), lacking cross modal deep collaboration between music and dance;

The dynamic adjustment of art generation relies heavily on preset rules, and lacks real-time response and personalized adaptation capabilities to audience feedback[7].

2) Current Research Status in China

In recent years, domestic research has developed rapidly and conducted extensive explorations around intelligent performance and real-time interaction.

(1) In terms of audience feedback analysis, scholars have proposed multimodal feedback collection schemes, but they mostly focus on a single scene and lack a universal feedback analysis model;

(2) In terms of multi-agent applications, some studies have applied multi-agent technology to virtual dance choreography and music accompaniment generation, but the system's collaboration and real-time performance need to be improved;

(3) In terms of music and dance collaboration, attempts have been made to achieve music and dance adaptation through rule matching or simple machine learning models, but the dynamic adjustment ability and artistic expression flexibility are insufficient.

Prominent gaps in domestic research:

Lack of deep integration of real-time audience feedback and multi-agent collaboration technology, and a lack of overall system architecture design;

The research on cross modal collaboration mechanism is not deep enough, making it difficult to achieve dynamic and accurate matching between music and dance;

Empirical research is mostly limited to laboratory environments and lacks large-scale, long-term testing in real performance scenarios. The practicality and generalization ability of the system need to be verified[8-10].

3) Research Review

In summary, existing research has not yet formed a complete system of "real-time feedback perception intelligent collaborative generation dynamic optimization iteration", which is difficult to meet the core needs of contemporary intelligent performance for interactivity, collaboration, and personalization. This article aims to address the above-mentioned gaps by integrating real-time audience feedback analysis and multi-agent collaboration technology to develop a dynamic music and dance coordination multi-agent system. The system's effectiveness is verified through empirical testing in multiple scenarios, forming a systematic solution that combines theory and practice.

D. RESEARCH IDEAS, METHODS, AND INNOVATIVE POINTS

1) Research Approach

This article follows the technical route of "theoretical sorting system design technical implementation empirical verification optimization and improvement":

1. Systematically review the core technologies and theories in real-time feedback analysis, multi-agent collaboration, music and dance generation, and clarify the key points and feasibility of technology integration;
2. Design a multimodal audience feedback collection and analysis system, and construct an emotional intention recognition model for feedback information;
3. Build a multi-agent system architecture, design cross modal collaboration mechanisms and dynamic generation algorithms;
4. Develop a prototype of a multi-agent system, complete hardware integration and software debugging;
5. Select typical scenarios for empirical testing, collect data, and verify system effectiveness;

6. Based on test results, optimize system algorithms and functions, and improve technical solutions and practical guidelines.

2) Research Methods

(1) Literature research method: Systematically review research achievements in real-time feedback analysis, multi-agent systems, music and dance generation, reinforcement learning, and other fields, laying a theoretical and technical foundation;

(2) Requirement analysis method: Through questionnaire surveys, expert interviews, and other methods, clarify the needs of users for system functionality, performance, and artistic expression in different scenarios;

(3) System design method: adopting modular design ideas, constructing a multi-agent system architecture, designing the functions and interaction processes of each module;

(4) Technical implementation method: Based on tools such as Python, TensorFlow, Unity, etc., core functions such as feedback collection, model training, and collaborative generation are implemented;

(5) Empirical testing method: Select three typical scenarios, invite audience, artists, and technical experts to participate in system testing, and verify system effectiveness through quantitative data and qualitative evaluation;

(6) Statistical analysis method: Using tools such as SPSS 26.0 and Matlab, descriptive statistics, difference analysis, and correlation analysis were conducted on the test data to quantitatively evaluate the system performance.

II. CORE CONCEPTS AND THEORETICAL FOUNDATIONS

A. THEORETICAL BASIS

Real time feedback analysis theory is a general term for a series of theories and methods related to the collection, preprocessing, feature extraction, and intent recognition of multimodal feedback data. Its core viewpoints include:

The fusion analysis of multimodal feedback data can improve the accuracy of intention identification, and single modal data can not fully reflect the real needs of the audience;

The temporal characteristics of feedback data are crucial for emotional intention recognition, as they can capture the dynamic evolution of emotional states;

The feedback analysis model needs to have real-time and robustness, and be able to adapt to individual differences in different environments and users[11].

This theory provides direct guidance for the design of a multimodal audience feedback collection system and the construction of an emotional intention recognition model in this article[12].

The theory of multi-agent collaboration originates from the field of distributed artificial intelligence, and its core research focuses on how multiple agents can achieve complex tasks through interactive collaboration. Its core viewpoints include:

The collaborative mechanism is the core of multi-agent systems, including various forms such as contract network protocols, reinforcement learning collaboration, game theory collaboration, etc;

The information sharing and dynamic negotiation between intelligent agents can improve the overall performance of the system and achieve a synergistic effect of "1+1>2";

Hierarchical collaborative architecture can reduce system complexity and improve system flexibility and scalability[13-14].

This theory provides theoretical support for the design of multi-agent system architecture and the development of cross modal collaboration mechanisms in this article.

Reinforcement learning theory is an important branch of machine learning that studies how agents interact with their environment and learn optimal behavioral strategies through trial and error learning. Its core elements include agents, environments, states, actions, and rewards, and policy optimization is achieved through Markov decision processes (MDP).

This theory provides a core approach for the design of cross modal collaboration mechanisms for multiagent systems in this article. By defining reasonable reward functions (such as music and dance collaboration matching degree, audience feedback satisfaction), it drives each agent to autonomously learn the optimal collaboration strategy[15].

The theory of music and dance generation covers music creation theory, dance choreography theory, and intelligent generation technology. The core research is how to achieve automatic generation of artistic content through algorithms. Its core viewpoints include:

Music generation needs to follow music theory rules such as rhythm, melody, and harmony, while also considering emotional expression and style consistency;

Dance generation needs to consider human kinematic constraints, motion coherence, and aesthetic principles of formation;

The key to cross modal generation lies in establishing a mapping relationship between music and dance, achieving collaborative unity in artistic expression[16].

This theory provides both artistic and technical guidance for the algorithm design of music generation agents and dance choreography agents in this article.

The theory of affective computing was proposed by Picard, which studies how to recognize, express, and simulate human emotions through computer technology. Its core viewpoints include:

Emotion is the core element of human interaction, and intelligent systems should have the ability to perceive and respond to emotions;

Multimodal physiological and behavioral data is the core basis for emotion recognition, which can effectively reflect human emotional states;

The goal of emotional computing is to achieve emotional interaction between humans and systems, and to enhance the intelligence and humanization level of the system.

This theory provides the core theoretical basis for the construction of the audience feedback emotion intention recognition model in this article[17].

These theoretical pillars are integrated into a cohesive functional pipeline: Affective Computing and Real-time Feedback Theory govern the Perception Layer for multi-modal data fusion; these outputs serve as the State (S) input for the Multi-Agent Collaboration framework. Within the Agent Layer, Music/Dance Generation Theories define the action constraints, while Reinforcement Learning acts as the Optimization Engine, dynamically mapping audience emotional states to cross-modal artistic parameters to achieve systemic equilibrium.

III. DESIGN OF REAL TIME AUDIENCE FEEDBACK COLLECTION AND ANALYSIS SYSTEM

A. DESIGN PRINCIPLES

1) Multimodal Fusion Principle

Comprehensive collection of physiological, behavioral, subjective evaluation and other multimodal feedback data to avoid the limitations of single modal data and improve the accuracy and comprehensiveness of emotional intention recognition[18-20].

2) Real Time Principle

Optimize the data collection and transmission process to ensure that the collection, preprocessing, and analysis of feedback data are completed within seconds, meeting the real-time response requirements of the system.

3) Non Invasive Principle

Adopting non-contact and low interference collection devices and methods to avoid interference with the viewing experience of the audience and improve user acceptance.

4) Principle of Robustness

The collected data and analysis results should directly serve the collaborative generation of music and dance, avoiding redundant data collection and reducing system computational load.

B. MULTIMODAL AUDIENCE FEEDBACK COLLECTION SYSTEM

1. Physiological data collection: A hybrid sampling approach is adopted to ensure feasibility. High-precision wearable devices (smart bracelets) are deployed to a representative “Core Sample Group” (n=50), while the remaining audience (n=450) is monitored via non-contact infrared thermal imaging for heart rate estimation ; Behavioral data collection: General audience engagement is captured via high-definition camera arrays for facial and skeletal analysis, while eye-tracking is limited to a subset of fixed-seat “Pilot Participants” to reduce logistical complexity and hardware costs;

2. Behavioral data collection: High definition cameras (frame rate $\geq$ 30fps) are used to collect facial expressions and body movement data, and eye tracking data is collected through eye tracking devices;

3. Subjective data collection: Develop a multi terminal interactive interface (mobile mini program, on-site touch screen, online interactive window), supporting real-time audience rating and text input; Use a microphone array to collect voice feedback data.

4. Preprocessing stage: Before the performance begins, guide the audience to complete device connections (such as binding smart bracelets, authorizing camera access), and complete baseline data collection (such as resting heart rate, skin conductivity);

5. Real time collection stage: During the performance, physiological and behavioral data is synchronously collected at fixed time intervals (500ms/time), and the audience can submit subjective feedback at any time;

6. Data transmission stage: 5G/WiFi6 technology is used to achieve low delay data transmission, and preliminary data pre-processing (denoising, format conversion) is performed through edge computing nodes;

7. Data storage stage: Real time feedback data is stored using a time-series database (InfluxDB), which supports high concurrency writing and fast queries.

C. AUDIENCE FEEDBACK ANALYSIS MODEL DESIGN

1) Data Preprocessing

1. Physiological data preprocessing: Using sliding average filtering to remove noise from heart rate and skin conductivity data, and normalizing the data to the [0,1] interval to eliminate individual differences;

2. Behavioral data preprocessing: Frame extraction, face detection, and pose estimation are performed on video data using the MediaPipe framework. Facial keypoints (68) and human pose keypoints (33) are extracted to generate action feature vectors;

3. Subjective data preprocessing: segment text comments, remove stop words, construct sentiment dictionaries, denoise speech data, extract features (MFCC coefficients), and standardize rating data.

2) Feature Extraction

1. Physiological characteristics: Extract 12 dimensional features such as mean heart rate, heart rate variability, peak skin conductivity, and facial temperature change rate;

2. Behavioral features: Extract facial expression features (based on VGGFace2 model), limb action features (based on LSTM to extract temporal action features), gaze concentration and other 18 dimensional features;

3. Subjective features: Extract 8-dimensional features such as text comment sentiment polarity, speech sentiment category, and average rating;

4. Fusion features: By using attention mechanism to weight and fuse multimodal features, a 38 dimensional audience feedback fusion feature vector is generated, highlighting features that contribute highly to emotional intention recognition.

3) Emotional Intent Recognition Model

Construct a multimodal emotion intention recognition model based on BERT+BiGRU, with the following model architecture:

1. Input layer: Input multimodal fusion feature vectors;
2. Encoding layer: The BERT model is used for semantic encoding of textual features (text comments, speech emotion text conversion), and the BiGRU model is used for temporal encoding of temporal features (heart rate changes, action sequences);
3. Fusion layer: dynamically adjust the weights of different modal encoding features through attention mechanism to achieve deep feature fusion;
4. Output layer: Using a fully connected layer and Softmax activation function, output the audience's emotional intention classification results (pleasure, excitement, calmness, boredom, dissatisfaction) and confidence level.

To resolve conflicting feedback from a large-scale audience, we implement a Weighted Consensus Mechanism. The system calculates a Global Sentiment Vector $V_g = \sum_i(\omega_i c_i)$, where ω_i represents the confidence level of the i-th agent and c_i is the sentiment category. When a bimodal distribution occurs (e.g., a split between "Boredom" and "Excitement"), the Collaboration Control Agent prioritizes the "dissatisfaction" or "boredom" signals to trigger a transition, ensuring the performance remains engaging for the lowest-common-denominator sentiment group.

During the model training process, the cross entropy loss function is used to optimize parameters through the Adam optimizer, and the early stopping method is introduced to prevent overfitting.

4) Feedback Analysis Model Performance Evaluation

5) Dataset Construction

Collect feedback data from 500 audiences of different ages, genders, and professions in three performance scenarios, and construct a labeled dataset containing 1.2 million multimodal feedback data, with the training set accounting for 70%, validation set accounting for 20%, and testing set accounting for 10%.

Evaluation indicators

Accuracy, Precision, Recall, and F1 Score are used as performance evaluation metrics for the model.

6) Evaluation Results(Table1)

Result analysis: The model has a high accuracy in identifying strong emotional intentions such as pleasure and excitement (all exceeding 93%), and a relatively low accuracy in identifying weak emotional intentions such as boredom and dissatisfaction. However, the overall average accuracy is 92.3%, and the F1 score is 91.8%, which can meet the system's demand for accurate recognition of audience emotional intentions.

IV. DESIGN OF MULTI AGENT SYSTEM ARCHITECTURE FOR DYNAMIC MUSIC AND DANCE COORDINATION

A. OVERALL SYSTEM ARCHITECTURE

Based on the theory of multi-agent collaboration and reinforcement learning, a "four layer architecture+four types of intelligent agents" multi-agent system architecture is constructed, including the perception layer, intelligent agent layer, collaborative control layer, and application layer. The specific architecture is as follows table 2.

B. INTELLIGENT AGENT DESIGN

1) Feedback Analysis Agent

1. Core function: Receive multimodal feedback data from the perception layer, call the emotion intention recognition model, and output the audience's emotion intention classification results and confidence level; Analyze the temporal trend of feedback data and predict the direction of audience emotional evolution;

2. Key modules: data receiving module, intent recognition module, trend prediction module, result output module;

3. Interactive interface: Receive preprocessed data from the perception layer and output emotional intention analysis results and trend prediction information to the collaborative control agent.

2) MusicGenerationAgent

1. Core function: Based on the instructions of collaborative control agents, generate music content that is adapted to the audience's emotional intentions and performance scenes; Support real-time adjustment of music rhythm, melody, instrument, speed and other parameters;

2. Key modules: parameter configuration module, melody generation module, rhythm generation module, instrument arrangement module, music synthesis module.

3) DanceChoreography Agent

1. Core function: Based on instructions from collaborative control agents, generate dance movements and formations that are adapted to music content and audience emotional intentions; Support real-time adjustment of dance style, movement amplitude, formation change frequency and other parameters;

2. Key modules: action library management module, action generation module, formation arrangement module, kinematic constraint module, dance rendering module.

4) Collaborative Control Agent

Key modules: instruction parsing module, collaborative strategy formulation module, conflict coordination module, instruction issuance module;

Technical implementation: The agent employs a Deep Q-Network (DQN) for strategy optimization. We define

TABLE 1. Performance of Audience Feedback Analysis Model

Emotional Intention Category	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Pleasure	94.6	93.8	95.2	94.5
Excitement	93.8	92.7	94.3	93.5
Calmness	91.5	90.3	92.1	91.2
Boredom	89.7	88.5	90.2	89.3
Dissatisfaction	90.3	89.6	91.1	90.3
Average	92.3	91.0	92.6	91.8

TABLE 2. Overall Architecture of Multi-Agent System

Architecture Layer	Core Function	Key Components	Technical Support
Perception Layer	Multimodal audience feedback collection, data preprocessing, feature extraction	Feedback collection devices, data transmission module, preprocessing module	5G/Wi-Fi 6, edge computing, MediaPipe
Agent Layer	Feedback analysis, music generation, dance choreography, collaborative decision-making	Feedback Analysis Agent, Music Generation Agent, Dance Choreography Agent, Collaboration Control Agent	Deep learning, reinforcement learning, music and dance generation algorithms
Collaborative Control Layer	Agent interaction protocol, cross-modal collaboration mechanism, conflict coordination	Interaction protocol module, collaborative scheduling module, conflict resolution module	Distributed collaboration algorithms, reinforcement learning collaboration mechanisms
Application Layer	Performance display, user interaction, system management, data visualization	Virtual performance engine, interactive interface, system management platform, visualization module	Unity, WebGL, big data visualization technology

the State Space (S) as a vector comprising the audience emotional intensity (E), current music BPM (B), and dance motion amplitude (A). The Action Space (A) consists of discrete adjustment vectors $\Delta = \{\delta_{\text{music}}, \delta_{\text{dance}}\}$, where δ represents incremental changes to rhythm or movement intensity. Through the reward function R, the agent learns the optimal policy $\pi(s)$ to maximize long-term audience satisfaction:

Basic reward: Collaborative matching degree of music and dance (calculated based on the matching degree of rhythm, emotion, and style);

Feedback reward: audience feedback on the compatibility between emotional intention and artistic expression; To avoid logical circularity and instability, the reward signal is calculated as a "Differential Satisfaction Score." Instead of absolute satisfaction, the agent evaluates the change in sentiment (ΔE) following a specific artistic adjustment. If an adjustment in rhythm leads to a statistically significant positive shift in the audience's arousal levels compared to the previous state, a positive reward is issued.

Punishment items: system response delay, violation of motion kinematics, inconsistent music style. Through reinforcement learning training, collaborative control agents can autonomously learn the optimal collaborative strategy, achieving dynamic adaptation between music and dance.

C. CROSS MODAL COLLABORATIVE MECHANISM DESIGN

1) Collaborative Process

1. Perception trigger: The feedback analysis intelligent agent outputs real-time results of audience emotional intention analysis. When the confidence level of emotional intention is $\geq 85\%$ and the duration is ≥ 3 seconds, the collaborative adjustment process is triggered;

2. Instruction generation: The collaborative control intelligent agent generates music adjustment instructions (such as increasing rhythm and melody pitch) and dance adjustment instructions (such as increasing action amplitude and intensifying formation) based on the emotional intention type and evolution trend, combined with the current music and dance state;

3. Parallel execution: The music generation agent and dance choreography agent execute instructions in parallel to generate new music segments and dance sequences;

4. Collaborative verification: The collaborative control intelligent agent verifies the collaborative matching degree between newly generated music and dance. If the matching degree is $\geq 85\%$, an update is executed; If the matching degree is less than 85%, adjust the instruction again and generate it;

5. Dynamic update: The performance system updates music and dance content in real-time, completes a collaborative adjustment, and the entire process takes $\leq$ 200ms.

2) Collaborative Matching Indicator System

Build a music and dance collaborative matching index system consisting of 3 primary indicators and 8 secondary indicators to quantify the collaborative effect(see table 3).

Collaborative matching degree=$\sum$ (secondary indicator score $\times$ weight), with a maximum score of 100 points.

D. IMPLEMENTATION OF SYSTEM CORE TECHNOLOGIES

1) Music Generation Technology

1. GAN based melody generation: Using the MusicGAN model, input emotional intention parameters. Rather than simple major/minor mapping, the system utilizes a "Multi-dimensional Timbre-Mode Matrix." Pleasure can be mapped

TABLE 3. Music-Dance Collaboration Matching Index System

First-Level Indicator	Second-Level Indicator	Indicator Definition	Weight
Rhythm Matching	Beat Synchronization Rate	Degree of synchronization between dance action beats and music beats	0.25
Rhythm Matching	Speed Adaptation Degree	Degree of adaptation between dance action speed and music speed	0.15
Emotional Adaptation	Emotional Tone Consistency	Degree of consistency between musical emotion and dance emotion	0.20
Emotional Adaptation	Expression Intensity Matching Degree	Degree of matching between musical emotional expression intensity and dance action amplitude	0.10
Style Unification	Style Type Consistency	Degree of consistency between musical style and dance style	0.15
Style Unification	Element Echo Degree	Degree of echo between musical melody transitions and dance formation changes	0.08
Style Unification	Instrument-Action Matching Degree	Degree of matching between specific instrument timbre and dance action characteristics	0.05
Overall Coordination	Overall Coordination	Degree of coordination and aesthetics presented by music and dance as a whole	0.02

to major modes, but “Excitement” or “Complex Pleasure” may trigger Dorian or Lydian modes combined with syncopated polyrhythms, allowing for cultural nuance and preventing artistic rigidity;

2. Dynamic rhythm adjustment: Based on reinforcement learning training, the rhythm adjustment model adjusts the beat speed (BPM) and rhythm type (such as segmented rhythm and evenly distributed rhythm) in real time according to the audience’s emotional intention;

3. Instrument arrangement optimization: Build an emotional mapping library for instrument timbre, and automatically select suitable instruments based on emotional intentions (such as drums and electric guitars for excitement, and piano and violin for calmness).

2) Dance Generation Techniques

1. Action sequence generation: Using a variational autoencoder (VAE) based action generation model, basic actions are extracted from the action library to generate action sequences that conform to emotional intent and kinematic constraints;

2. Action music synchronization: Using the Dynamic Time Warping (DTW) algorithm, precise synchronization between dance movements and music beats is achieved with a synchronization error of $\leq$ 50ms.

3) Real Time Response Optimization

1. Distributed computing architecture: the distributed architecture of edge computing+cloud collaboration is adopted. Tasks with high real-time requirements such as feedback analysis and collaborative decisionmaking are executed at edge nodes, and computing intensive tasks such as music and dance generation are executed on the cloud;

2. Model lightweighting: Pruning and optimizing deep learning models to reduce computational complexity and improve inference speed;

3. Pre generation caching mechanism: Pre generate music and dance action clips with different emotional intentions and styles, store them in the cache, and directly call and quickly adapt them during collaborative adjustment to shorten response time.

V. EMPIRICAL TESTING DESIGN AND RESULT ANALYSIS

A. EMPIRICAL TEST DESIGN

1) Test Scenarios

Select three typical performance scenarios for empirical testing:

1. Virtual Idol Concert Scene: Combining online and offline, virtual idols perform music and dance, and audiences submit feedback through mobile apps and live interactive screens;

2. Interactive Art Exhibition Scene: Immersive Art Space, where the audience interacts with the system through body movements, physiological devices, and generates adapted music and dance performances;

3. Online entertainment live streaming scene: The anchor and virtual characters perform in conjunction, and the audience participates in interaction through bullet comments, gifts, and voice feedback. The system dynamically adjusts the performance content.

2) Test Object

Invite 800 test participants (600 audience members, 100 artists, 100 technical experts), including audiences of different ages (1855 years old), genders (320 males, 280 females), and professions (students, professionals, freelancers, etc.); Artists include music producers, dance choreographers, and digital art creators; Technical experts include researchers and engineers in the fields of artificial intelligence, intelligent interaction, and digital media.

3) Testing Equipment and Environment

1. Hardware equipment: high-performance server (CPUIntelXeonPlatinum8375C, GPUNVIDIAA100), edge computing node, high-definition camera, smart bracelet, eye tracker, interactive screen, audio system, VR equipment (some scenes);

2. Software environment: Ubuntu 20.04 operating system Python3.9, TensorFlow2.10, PyTorch1.13, Unity2022, InfluxDB2.7.

4) Test Indicators

1. System performance indicators: feedback recognition accuracy, music and dance collaborative matching degree,

system response delay, concurrent processing capability;

2. User experience indicators: audience satisfaction (using Likert 10 point scale), active participation (proportion of audience participating in feedback), emotional resonance (degree of fit between audience emotions and performance content);

3. Artistic expression indicators: artistic style adaptability, emotional expression accuracy, and creative flexibility (artist evaluation);

4. Scenario adaptation indicators: system stability, functional adaptability, and user acceptance in different scenarios.

5) Testing Process

1. Pre testing phase (1 month): Conduct small-scale testing (100 viewers) in a single scenario to optimize system performance and interactive experience;

2. Formal testing phase (4 months): Conduct large-scale testing in 3 types of scenarios, with 8 tests per scenario, each test lasting 90 minutes, and real-time collection of system performance data and user feedback data;

3. Post testing phase (1 month): Collect feedback and suggestions from test participants through questionnaire surveys, in-depth interviews, expert reviews, and other methods, and conduct comprehensive analysis.

B. QUALITATIVE FEEDBACK ANALYSIS

1. Audience interview feedback:

87.5% of the audience believe that the system can accurately capture my emotions and adjust the performance content to meet my expectations;

82.3% of the audience stated that “real-time interaction makes me more engaged and no longer passively watching”;

78.6% of the audience hopes that the system can provide more personalized settings, such as custom performance styles.

2. Artist interview feedback:

90.0% of artists believe that the system provides new ideas and tools for artistic creation, greatly improving creative efficiency;

85.7% of artists stated that ‘real-time feedback data helps understand audience preferences and optimize artistic expression’;

83.3% of artists suggest adding more art style templates to enhance the diversity of artistic creation in the system.

3. Technical expert review feedback:

92.0% of experts believe that the system architecture design is scientifically reasonable and the multi-agent collaboration mechanism is efficient;

88.0% of experts stated that the feedback parsing model and collaborative generation algorithm are innovative and practical;

85.0% of experts suggest “further optimizing the robustness of the system in complex environments and expanding more application scenarios”.

C. EMPIRICAL CONCLUSION

The empirical test results indicate that:

The system has excellent artistic expression ability and can accurately transform audience feedback into artistic expression, providing effective creative assistance for artists and receiving high recognition from artists; The system has demonstrated good adaptability in scenarios such as virtual idol concerts, interactive art exhibitions, and online entertainment live broadcasts, and has broad promotional and application value; There is still some room for optimization in the system, such as insufficient personalized settings, limited artistic style templates, and the need to improve robustness in complex environments, which need to be further improved in subsequent research.

VI. SYSTEM OPTIMIZATION AND PROMOTION SUGGESTIONS

A. SYSTEM OPTIMIZATION STRATEGY

1) Function Optimization

Artistic Style Expansion: Expand the music style library (adding ethnic, rock, folk, and other styles) and dance style library (adding ethnic, modern, and street dance styles) to increase artistic diversity;

Multi role collaborative optimization: supports collaborative performances of multiple virtual characters (such as 5-10 dance characters), optimizes multi character formation arrangement algorithms, and enhances the coordination and viewing of group dances.

2) Performance Optimization

Response speed optimization: Optimize the distributed computing architecture and adjust the task allocation ratio between edge nodes and the cloud; Expand the pre generated cache to cover more emotional intentions and style combinations, further reducing response latency.

3) Interaction Optimization

1. Simplification of interactive interface: Optimize the design of multi terminal interactive interface, reduce operational complexity, and improve the convenience of use for middle-aged and elderly audiences;

2. Rich feedback methods: adding non-contact feedback methods such as gesture recognition and facial expression control to reduce dependence on devices;

3. Feedback visualization: Increase the real-time visualization function of audience feedback, allowing the audience to intuitively understand the impact of their feedback on the performance content and enhance their sense of participation.

B. PROMOTION SUGGESTIONS

1. Pilot promotion stage (12 years): Select industry-leading enterprises to carry out pilot cooperation, verify the commercial value and application effect of the system in core scenarios (such as virtual idol concerts), and accumulate successful cases;

2. Regional promotion stage (23 years): Based on successful pilot projects, promote to relevant enterprises in specific regions (such as first tier cities and areas with developed cultural industries) and establish regional cooperation networks;

3. National promotion stage (35 years): Summarize promotion experience, improve system functions and service system, promote application nationwide, and form a large-scale commercial landing.

1) Suggestions for Ecological Construction

1. Technical Ecological Cooperation: Establish technical cooperation relationships with enterprises in the fields of artificial intelligence, intelligent hardware, cloud computing, etc., jointly build a technical ecosystem, and share resources and achievements;

2. Co construction of Content Ecology: Invite artists and creators to participate in the construction of the system's content library (music, dance, style templates), forming an open content ecology;

3. Standard system construction: Collaborate with industry associations and research institutions to develop technical standards, data standards, and evaluation standards for real-time interactive intelligent performance systems, and promote the standardized development of the industry;

4. Talent cultivation: Cooperate with universities and vocational colleges to offer majors in intelligent art creation, digital entertainment technology, and other related fields, cultivate composite talents with both artistic literacy and technical abilities, and provide talent support for industry development.

VII. RESEARCH CONCLUSION

This article focuses on the development of a dynamic music and dance coordination multi-agent system driven by real-time audience feedback, specifically designed to simulate the rhythmic mechanical movements and complex fiber morphology inherent in the industry. By integrating these industrial parameters, the system achieves a precise multimodal synergy that reflects the spatial distribution and technical excellence of modern engineering. Through theoretical analysis, system design, technical implementation, and empirical testing, the following conclusions are drawn:

Real time audience feedback contains rich emotional intentions and aesthetic preferences, and is the core resource driving dynamic optimization of music and dance performances. Building a "physiological behavior subjective" multimodal feedback collection system can comprehensively and accurately capture audience feedback information. Based on the BERT+BiGRU multimodal emotional intention recognition model, it can achieve accurate recognition of audience emotional intentions with an average accuracy of 92.3%, providing reliable decision-making basis for the system's real-time response.

AUTHOR CONTRIBUTIONS

Conceptualization –Yue Yu and Xi Zhang; methodology – Yue Yu and Xi Zhang; investigation –Yue Yu and Xi Zhang; writing-original draft preparation – Yue Yu and Xi Zhang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] D. Neumann, B. Bonenti, and K. Murphy, "Dynamic Musical Stimuli: Subjective and Physiological Responses to an Original Music Stimulus Set to Study Emotion," International Journal of Psychophysiology, vol. 213, no. S, pp. 113010–113010, 2025, doi: 10.1016/j.ijpsycho.2025.113010.

[2] A. Knutas, H. Hämäläinen, J. Ikonen, et al., "Implementation of an embedded mobile device based feedback system for real-time audience feedback," Computer Applications in Engineering Education, vol. 21, no. S1, pp. E31-E40, 2013, doi: 10.1002/cae.20570.

[3] J. Chen, R. Mak, and S. Fujita, "The Effect of Combination of Video Feedback and Audience Feedback on Social Anxiety," Behavior Modification, vol. 39, no. 5, pp. 721–739, 2015, doi: 10.1177/0145445515587087.

[4] Ask for Audience Feedback, "Nonprofit Communications Report," 2016; 14(4):6-6, doi: 10.1002/npcr.30391.

[5] J. Chen, J. McLean E, and E. Kemps, "Combined Audience and Video Feedback With Cognitive Review Improves State Anxiety and Self-Perceptions During Speech Tasks in Socially Anxious Individuals," Behavior Therapy, vol. 49, no. 2, pp. 165–179, 2018, doi: 10.1016/j.beth.2017.11.006.

[6] O. Ruvalcaba, J. Shulze, A. Kim, et al., "Women's Experiences in eSports: Gendered Differences in Peer and Spectator Feedback During Competitive Video Game Play," Journal of Sport & Social Issues, vol. 42, no. 4, pp. 295–311, 2018, doi: 10.1177/0193723518773287.

[7] E. Molleman, "How a powerful audience and audience feedback moderate the relationship between performance approach orientation and exaggerated self-promotion," The International Journal of Human Resource Management, vol. 30, no. 15, pp. 2273–2292, 2019, doi: 10.1080/09585192.2017.1322123.

[8] K. O. Aina, "Deep Reinforcement Learing for Multi-Agent Coordination," 2025, doi: 10.1007/s10015-025-01089-z.

[9] G. A, Y. Y, and U. K, "Feedback assessment from the audience as part of health literacy training for health professionals: a case from Fukushima after the nuclear accident," Annals of the ICRP, vol. 50, no. 1_suppl, pp. 1466453211010919–1466453211010919, 2021, doi: 10.1177/01466453211010919.

[10] J. Freudendal, "Rationalising creative uncertainty? The negotiated use of audience feedback in Danish film production," Nordicom Review, vol. 46, no. 2, pp. 250–272, 2025, doi: 10.2478/nor-2025-0020.

[11] P. Christos, G. Dimitris, and D. Maria, "Dynamic Musicscapes in Northern Greece: A Roma Case Study," Folklore: Electronic Journal of Folklore, vol. 74, pp. 97–114, 2018, doi: 10.7592/fejf2018.74.greece.

[12] E. Zheng, G. Kondo Y, S. Zilora, et al., "Tag-aware dynamic music recommendation," Expert Systems With Applications, vol. 106, pp. 244–251, 2018, doi: 10.1016/j.eswa.2018.04.014.

[13] Y. Kangyou, W. Shaoyue, X. Wenge, et al., "Cheer for me: effect of non-player character audience feedback on older adult users of virtual reality exergames," Virtual reality, vol. 27, no. 3, pp. 11–17, 2023, doi: 10.1007/s10055-023-00780-5.

[14] P. Ashlee, G. Sarah, and G. John, "Instrumented measures of postural control in college dancers with and without chronic ankle instability," Physical Therapy in Sport, vol. 58, pp. 41–45, 2022, doi: 10.1016/j.ptsp.2022.09.001.

[15] R. Paola, “The Implantation of a Dance Workshop on the Quality of Life in the Work Environment: The Paola Rettore Method,” International Journal of Art, Culture, Design, and Technology (IJACDT), vol. 11, no. 2, pp. 1–9, 2022, doi: 10.4018/IJACDT.305799.
[16] M. Fouad, “Dance and the Twilight of Capitalism,” Dance Chronicle, vol. 45, no. 3, pp. 263–266, 2022, doi: 10.1080/01472526.2022.2105086.
[17] F. Barrett S and P. Janata, “Neural responses to nostalgia-evoking music modeled by elements of dynamic musical structure and individual differences in affective traits,” Neuropsychologia, vol. 91, pp. 234–246, 2016, doi: 10.1016/j.neuropsychologia.2016.08.012.
[18] R. Jonas, ““My Walking is My Dancing”: The Relationship between Dance and Music in Anne Teresa De Keersmaeker’s Work,” Dance Chronicle, vol. 45, no. 3, pp. 229–249, 2022, doi: 10.1080/01472526.2022.2108278.
[19] H. Colleen, “Accounting and Accountability: How the NEA Funded Dance,” Dance Chronicle, vol. 45, no. 3, pp. 250–253, 2022, doi: 10.1080/01472526.2022.2110633.
[20] J. He, “Analyzing film and drama reviews: Distinguishing trolls from genuine audience feedback based on the BERT model,” in Proceedings of the 2024 International Conference on Applied Statistics, Modeling and Advanced Algorithms. 2024, doi: 10.54254/2755-2721/53/20241376.