Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 2 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i02.60139

PAPER

A Multi-Sensor Fusion Approach for Music Rhythm-Based Rehabilitation System with Mobile Interaction

Yang Liu[1] (✉), **Xiaoming Yi**[2]

[1]Henan Quality Institute, Pingdingshan, China

[2]Tieling Normal College, Tieling, China

liuyang821031@126.com

ABSTRACT

Motor dysfunctions caused by diseases such as stroke and Parkinson's disease present significant challenges for traditional home-based rehabilitation, including low training adherence and a lack of quantitative assessment. The proliferation of mobile devices and the advancement of multi-sensor technologies offer new pathways for personalized rehabilitation. Music rhythm can promote neural plasticity through the auditory-motor neural pathways, but its application is hindered by issues such as interference from single-sensor data, insufficient robustness of traditional fusion algorithms, lack of personalized rhythm adaptation, and imbalances in accuracy and real-time performance on mobile platforms. To address these challenges, this paper proposes a multi-sensor fusion mobile interaction music rhythm rehabilitation system. The core contributions are: 1) Developing a multi-sensor architecture comprising "Inertial Measurement Unit (IMU) + portable Electromyography (EMG) + built-in microphone" to achieve three-dimensional data collection of "motion trajectory-muscle activation-rhythm synchronization," balancing portability and completeness; 2) Designing a lightweight algorithm chain that processes data through Kalman filtering and wavelet denoising, using convolutional neural network – long short-term memory (CNN-LSTM) for end-to-end fusion of multi-modal temporal features, combined with proximal policy optimization – generative adversarial network (PPO-GAN) for generating adaptive rhythms, overcoming mobile device resource constraints; 3) Conducting case-control experiments to establish a quantitative rehabilitation assessment system. This system provides an integrated solution for home rehabilitation, offering "precise perception-intelligent adaptation-quantitative evaluation," and presents a new paradigm for interdisciplinary research in mobile healthcare and neurorehabilitation.

KEYWORDS

multi-sensor fusion, mobile interaction, music rhythm rehabilitation, convolutional neural network – long short-term memory (CNN-LSTM), reinforcement learning, motor function recovery

Liu, Y., Yi, X. (2026). A Multi-Sensor Fusion Approach for Music Rhythm-Based Rehabilitation System with Mobile Interaction. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(2), pp. 116–130. https://doi.org/10.3991/ijim.v20i02.60139

Article submitted 2025-07-24. Revision uploaded 2025-11-29. Final acceptance 2025-12-16.

1 INTRODUCTION

Neurodegenerative diseases such as stroke and Parkinson's disease have become the leading causes of motor dysfunction worldwide. According to the World Health Organization, over 15 million new stroke patients are added each year globally, with approximately 70% of these patients suffering from motor dysfunction, which severely impacts their quality of life and creates a significant social and healthcare burden [1–3]. Traditional rehabilitation training highly depends on professional hospital equipment and medical guidance, which results in issues such as limited training scenarios and monotonous procedures. In contrast, home rehabilitation faces challenges such as the lack of a scientifically quantifiable assessment system and personalized guidance, leading to patient adherence rates generally below 60%, severely restricting rehabilitation outcomes [4, 5]. Against this backdrop, exploring low-cost, highly portable, and effective home rehabilitation solutions has become a research hotspot and an urgent need in the field of neurorehabilitation.

The global penetration rate of mobile intelligent devices has exceeded 85% [6–8]. The built-in Inertial Measurement Unit (IMU), highly sensitive microphones, and other sensor modules, combined with external portable Electromyography (EMG) sensors, can achieve low-cost, non-invasive collection of motion states and physiological signals, providing ideal hardware support for home rehabilitation. At the same time, neuroscience research has confirmed that music rhythm can activate the neural pathways between the auditory cortex and motor cortex, regulating the synchronized firing of motor neurons, promoting muscle coordination and promoting neural remodeling. Its intervention effects in motor function rehabilitation have been verified by multiple clinical studies [9, 10]. The combination of mobile interactive technology and music rhythm rehabilitation mechanisms provides a new technological pathway to break through the limitations of traditional rehabilitation models.

Although relevant research has made some progress, existing solutions still have three major gaps: in the sensor fusion layer, traditional fusion algorithms struggle to handle the nonlinear characteristics of multimodal data and environmental noise interference. Most studies use single or dual sensor combinations, resulting in insufficient data dimensions that fail to comprehensively reflect the complete rehabilitation state of "motion trajectory-muscle activation-rhythm synchronization" [11, 12]; in the rhythm interaction layer, existing systems are mostly designed based on fixed rhythm templates, lacking dynamic adaptation to the patient's real-time motion state, making it difficult to balance the specificity and safety of rehabilitation training [13, 14]; in the mobile implementation layer, the contradiction between complex algorithms and device resource constraints is prominent. Current systems generally suffer from high real-time response delays and insufficient long-term operational stability. Additionally, they neglect the "perception-interaction-assessment" closed-loop design, resulting in single-mode feedback that fails to maintain long-term patient adherence [15, 16].

To address the above issues, this paper proposes a multi-sensor fusion-based mobile interaction music rhythm rehabilitation system. The research objectives are: 1) to design a multi-sensor fusion architecture adapted to mobile home environments, achieving precise and comprehensive perception of the patient's motion state and rhythm synchronization; 2) to optimize a lightweight algorithm chain to

achieve real-time fusion of multimodal data and personalized rhythm generation under mobile resource constraints; and 3) to develop a system prototype and verify its technical feasibility and clinical rehabilitation effectiveness through clinical experiments. The subsequent chapters are arranged as follows: Chapter 2 details the overall system architecture and core algorithm principles; Chapter 3 quantitatively analyzes the system's technical performance and rehabilitation effects through comparative experiments and clinical validation; Chapter 4 discusses the core significance of the experimental results, differences from existing research, and research limitations; and Chapter 5 concludes the paper and looks ahead to future research directions.

2 SYSTEM DESIGN AND ALGORITHM PRINCIPLES

2.1 Overall system architecture

The mobile interaction-based music rhythm rehabilitation system proposed in this paper adopts a bottom-up five-layer hierarchical architecture, where each layer is functionally independent yet deeply coupled, ensuring accurate data collection, efficient processing, and real-time interaction in mobile scenarios. Figure 1 shows the hierarchical architecture of the multi-sensor fusion-based mobile interaction music rhythm rehabilitation system. The sensor layer, as the core of data collection, integrates the IMU built into mobile devices, the external portable EMG sensor, and the built-in microphone: The IMU collects three-dimensional acceleration and angular velocity data at a sampling rate of 100 Hz, accurately capturing the patient's limb motion trajectory; the EMG patch is applied to the target muscle group and collects muscle activation signals at a sampling rate of 1000 Hz, reflecting the quality of force exertion and the neuromuscular control state; the built-in microphone synchronously collects music signals and the sound of the patient's movements, providing data support for rhythm synchronization evaluation. The three sensors work together to achieve a full coverage of three-dimensional data, including "motion trajectory-muscle activation-rhythm synchronization."

The preprocessing layer addresses noise and interference issues in multi-source data. Kalman filtering is applied to correct IMU data drift and improve posture estimation accuracy. The EMG signal is denoised using three-level decomposition and thresholding with the db4 wavelet basis. The short-time Fourier transform (STFT) is used to extract core features of the music signal, such as beat intensity and beats per minute (BPM), laying the foundation for subsequent fusion processing. The fusion algorithm layer employs the convolutional neural network – long short-term memory (CNN-LSTM) network as the core model, which end-to-end fuses the preprocessed multimodal temporal features and outputs the motion-rhythm matching degree and motion normalization score. The rhythm interaction layer dynamically generates personalized music rhythms based on the output of this model using the proximal policy optimization – generative adversarial network (PPO-GAN) cooperative model, enabling adaptive adaptation to the patient's real-time motion state. The evaluation and interaction layer, supported by a mobile app, provides rhythm visualization, real-time training feedback, rehabilitation index quantification, and historical data query functions, ensuring ease of use and traceability of training.

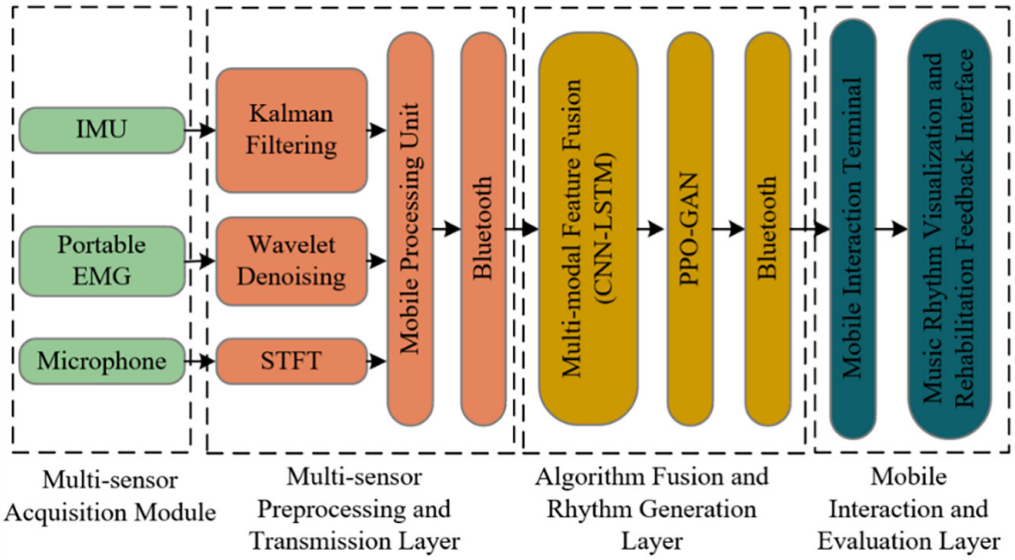

Fig. 1. Multi-sensor fusion-based mobile interaction music rhythm rehabilitation system architecture diagram

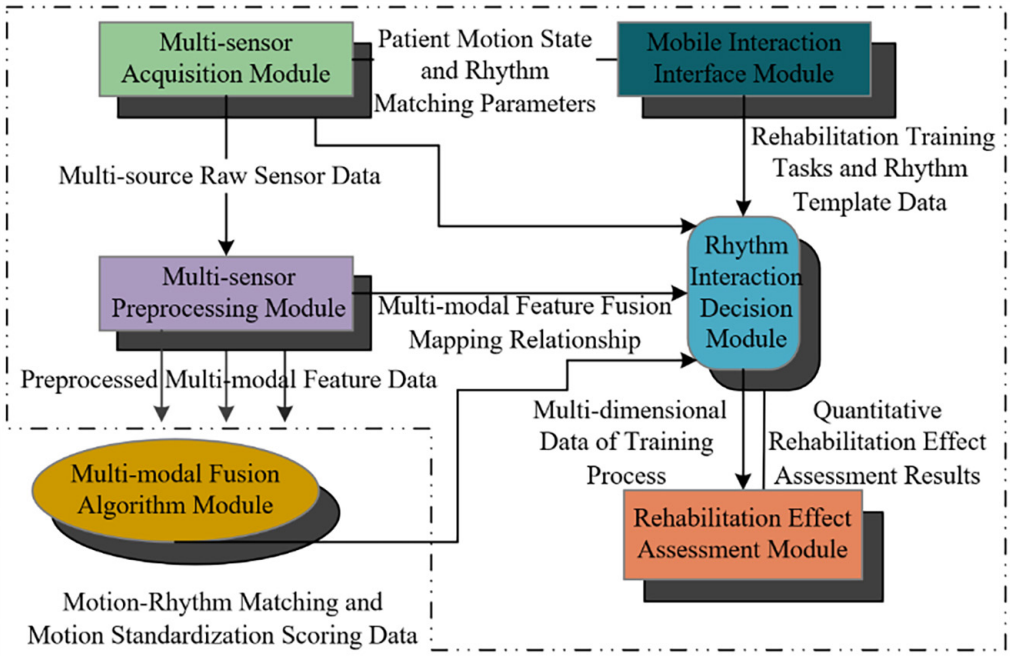

Fig. 2. Multi-sensor fusion-based mobile interaction music rhythm rehabilitation system data flow and evaluation architecture diagram

The system workflow forms a complete feedback loop: after the multi-sensor modules simultaneously collect data, time alignment is achieved through a soft synchronization mechanism based on signal correlation. After preprocessing and multi-modal feature fusion, the patient's motion state evaluation result is obtained, driving the rhythm interaction layer to generate an adaptive rhythm and provide feedback to the patient through the app. The patient adjusts their movements according to the music rhythm and evaluation prompts. The new motion state is then collected by the sensor layer and enters the next cycle, ensuring the specificity and continuity of the training. Figure 2 shows the entire process of multi-sensor data flow and rehabilitation evaluation architecture, from multi-source sensor data collection,

preprocessing, and fusion computation to rhythm interaction decision-making and quantifiable rehabilitation evaluation, clarifying the data transmission and functional cooperation between modules.

2.2 Multi-sensor preprocessing algorithm

The three-dimensional acceleration and angular velocity data collected by the IMU are prone to drift due to cumulative errors, which directly reduce the accuracy of motion trajectory and posture estimation. Therefore, the extended Kalman Filter (EKF) is used for data calibration. The core principle is to dynamically correct measurement biases through a "predict-update" iterative process, constructing state and observation equations to describe the system's dynamic characteristics: the state equation is defined as $x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}$, and the observation equation is $z_k = Hx_k + v_k$. Here, the state vector $x_k \in R^9$ includes 3-dimensional position, 3-dimensional velocity, and 3-dimensional posture angles, and the observation vector $z_k \in R^6$ corresponds to the IMU accelerometer and gyroscope measurements. A is the state transition matrix, H is the observation matrix, and w_{k-1} and v_k are process noise and observation noise, respectively, both following Gaussian distributions. The covariance matrices for these noises are calibrated using 10 sets of static and dynamic experimental data. The process noise covariance is $Q = \text{diag}(10^{-4}, 10^{-4}, 10^{-4}, 10^{-3}, 10^{-3}, 10^{-3}, 10^{-5}, 10^{-5}, 10^{-5})$, and the observation noise covariance is $R = \text{diag}(10^{-2}, 10^{-2}, 10^{-2}, 10^{-1}, 10^{-1}, 10^{-1})$. This method corrects the IMU drift in real time, ensuring that posture estimation errors are controlled within 2.5°, providing reliable data support for subsequent motion state evaluation.

EMG signals are typically non-stationary, weak signals that are susceptible to electromagnetic interference and myoelectric artifact contamination. Therefore, a 3-level wavelet decomposition with soft threshold denoising is performed using the db4 wavelet basis. The core reason for selecting the db4 wavelet basis is its compact support and orthogonality, which allows for the separation of noise while preserving the peak characteristics of the EMG signal. The specific processing procedure is as follows: First, the original EMG signal is subjected to 3-level wavelet decomposition to obtain one low-frequency approximation coefficient and three high-frequency detail coefficients. The high-frequency detail coefficients are processed using an improved soft threshold function $\omega_j' = \text{sign}(\omega_j)\left(|\omega_j| - \lambda\right)$, where $\lambda = \sigma(2\ln N)^{1/2}$, σ is the noise standard deviation, and N is the signal length. This suppresses the noise components. Finally, the pure EMG signal is reconstructed by combining the low-frequency coefficients with the modified high-frequency coefficients. Compared to traditional Butterworth filtering, this method improves the signal-to-noise ratio (SNR) of the EMG signal by over 15% and effectively preserves the amplitude and temporal characteristics of muscle force peaks, providing accurate neuromuscular activation data for motion quality assessment.

The rhythm features of the music signal are the core basis for rhythm interaction. STFT is used for feature extraction. Considering the non-stationary nature of the music signal, the STFT divides the signal into short-time stationary segments using a sliding window. A Hanning window of length 2048 points is selected, with a 50% window overlap to avoid feature loss. The Fourier transform is performed on each window to obtain the time-frequency matrix $S(t, f) = \text{STFT}(x(t))$, where t is the time axis and f is the frequency axis. The spectral peaks within the rhythm frequency range are calculated to determine the candidate positions for beats. The peak period is computed based on autocorrelation analysis and converted into BPM.

Meanwhile, the peak amplitude in the spectrum is extracted as the beat intensity feature, constructing a three-dimensional rhythm feature vector of "BPM-beat sequence-intensity" to provide a quantifiable basis for subsequent rhythm matching and adaptive generation.

2.3 Multi-modal feature fusion algorithm: CNN-LSTM

To achieve precise fusion of multi-modal temporal data and real-time operation on mobile devices, this paper designs a CNN-LSTM hybrid network architecture. The core advantage of this architecture is its ability to capture both local spatial features and long-term temporal dependencies while also being lightweight to fit within the resource constraints of mobile devices. The input layer uses a 100 ms sliding window to receive three types of modal data: IMU data after preprocessing, EMG data after denoising, and music rhythm feature vectors. These three types of data are dimensionally aligned and concatenated into a multi-modal input tensor. The CNN module is based on the lightweight MobileNet architecture and consists of three convolutional blocks, each containing a 3×3 depth-wise separable convolution layer, a batch normalization layer, and a ReLU activation function. The 3×3 convolution kernel is effective in extracting local spatial features, and the depth-wise separable convolution reduces the parameter computation by over 80% compared to traditional convolutions. The LSTM module uses a 2-layer bidirectional structure. The bidirectional design allows for the simultaneous capture of both forward and backward temporal dependencies of the data. The hidden layer dimensions are experimentally calibrated to balance feature expression ability and computational cost. The fusion layer introduces a self-attention mechanism by calculating attention weights $\alpha_i = \text{softmax}(W_a f_i + b_a)$, dynamically adjusting the contribution of IMU, EMG, and rhythm features. For example, when the signal-to-noise ratio of muscle activation is low, the weight of the EMG feature is automatically reduced. Here, W_a is the weight matrix, and f_i is the i-th modal feature. The output layer consists of two fully connected layers, which output the motion-rhythm matching degree and motion normalization score, directly serving rehabilitation assessment needs.

To ensure the fusion accuracy and generalization ability of the lightweight model, a targeted training strategy is designed. The dataset includes multi-sensor data from 20 healthy volunteers and 30 stroke rehabilitation patients, covering upper-limb/lower-limb rehabilitation movements. The data is labeled by two senior rehabilitation physicians using a 5-point scale based on a motion standard library and rhythm synchronization videos, annotating motion standard degree and rhythm matching degree. These annotations are linearly mapped to continuous labels ranging from 0 to 100, with an intraclass correlation coefficient (ICC) of 0.92, ensuring label reliability. The loss function is mean squared error (MSE), suited for regression tasks to predict scores, defined as:

$$\text{Loss} = \frac{1}{N}\sum_{i=1}^{N}[(y_{i1} - \hat{y}_{i1})^2 + (y_{i2} - \hat{y}_{i2})^2]$$

where, y_{i1}, y_{i2} are the true labels, $\hat{y}_{i1}$, $\hat{y}_{i2}$ are the model predictions, and N is the sample size. The optimizer used is Adam, with a learning rate of 0.001 and a weight decay rate of 0.9 to prevent overfitting.

In response to mobile resource constraints, knowledge distillation technology is employed to achieve model lightweighting: A complex model with "3-layer CNN + 3-layer bidirectional LSTM" serves as the teacher model, and its output soft labels are combined with real hard labels to train the lightweight student model.

The distillation temperature is set to 3 to balance the supervision of hard labels and the knowledge transfer of soft labels.

2.4 Adaptive rhythm generation algorithm: PPO + GAN

To achieve dynamic adaptation of rhythm parameters to the patient's real-time motion state, the PPO reinforcement learning algorithm is used to build the decision mechanism. Its core advantage is the use of clipping constraints in policy updates (clip) to avoid parameter mutations, which aligns with the core requirements of safety and stability in rehabilitation scenarios. The agent is defined as the rhythm adjustment module, with an output dimension of 2, consisting of two core parameters: one is the BPM adjustment amount, with a value range of ±5 BPM and a step size of 1 BPM, which avoids rhythm mutations that may lead to patient movement disorders; the other is the rhythm complexity level, corresponding to single beat, double beat, and multi-beat combinations, to adapt to the varying difficulty of movements required at different stages of rehabilitation. The state space is designed as a 3-dimensional continuous vector, including the action normalization score, the motion-rhythm matching degree, and the muscle fatigue level calculated based on the EMG signal, which are the outputs from the CNN-LSTM. The reward function is designed in a multi-objective weighted form, balancing rehabilitation effectiveness and safety constraints, defined as $R = \alpha \cdot M + \beta \cdot S - \gamma \cdot F$, where M is rhythm matching degree, S is motion standard degree, and F is muscle fatigue level. The weight coefficients are optimized through grid search combined with cross-validation, with the final values determined as $\alpha = 0.4$, $\beta = 0.5$, and $\gamma = 0.1$, prioritizing the motion standardization and rhythm synchronization while also suppressing excessive fatigue. During training, the policy network uses a 2-layer fully connected structure, and the value network shares feature extraction layers to reduce parameter redundancy. The clipping coefficient is set to 0.2 to ensure that the policy update magnitude remains controllable. After 5000 rounds of training, the policy converges, and the cumulative reward increases to 3.2 times the initial value, achieving a dynamic balance between "rehabilitation effectiveness" and "motion safety."

To address the auditory monotony of fixed rhythm templates, a GAN model based on LSTM-CNN is constructed to generate personalized rhythm sequences. The core objective is to generate rhythms that are both adaptable and natural within the rhythm parameters constrained by the PPO decision. The generator uses a 1-layer bidirectional LSTM structure, with input being a concatenated tensor of the 2D rhythm parameters from PPO output and a random noise vector. The output is a fixed-length rhythm sequence, where the sequence elements represent the mapping relationship between beat types and temporal positions. The discriminator uses a 3-layer CNN architecture, with input being the STFT time-frequency graph of the rhythm sequence. Through 3×3 convolution layers and max-pooling layers, the local timbre and temporal pattern features are extracted, and the output is a 0-1 authenticity score. The training adopts an alternating adversarial strategy: the generator aims to "minimize the discriminator's authenticity score error" while also introducing the PPO parameter constraint loss; the discriminator aims to "maximize the classification accuracy between real rhythms and generated rhythms." The training dataset includes rhythm sequences from 500 professional rehabilitation music tracks. After 100 rounds of adversarial training, the discriminator's authenticity score of the generated rhythm stabilizes above 0.75, and the patient's subjective auditory naturalness score reaches 4.3/5, achieving a balance between adaptability and auditory experience.

3 EXPERIMENTAL VALIDATION AND RESULT ANALYSIS

To comprehensively verify the technical performance and clinical effectiveness of the proposed multi-sensor fusion-based mobile interaction music rhythm rehabilitation system, a controlled experiment including both technical validation and clinical intervention was designed. The experimental process strictly follows the *Quality Management Standards for Clinical Research of Medical Devices*, and all participants signed informed consent forms. The experimental subjects were divided into two groups: the rehabilitation patient group included 30 stroke patients with motor dysfunction, consisting of 18 males and 12 females, aged 45–65 years, with an average age of 54.2 ± 6.7 years. The patients were classified into mild, moderate, and severe groups based on the Fugl-Meyer Assessment (FMA) score, excluding those with severe cognitive impairments or cardiovascular diseases. The healthy control group consisted of 20 healthy volunteers, including 12 males and eight females, aged 40–60 years, with an average age of 51.5 ± 7.3 years, with no history of motor dysfunction or neurological diseases.

The experimental devices included mobile terminals such as the iPhone 14 and iPad Pro; the sensor devices used were Shimmer3 wireless EMG sensors, and the Vicon Nexus 2.12 motion capture system was introduced as the gold standard for motion state evaluation. Two comparative systems were set up: the traditional fixed-rhythm rehabilitation app and the single-sensor fusion rehabilitation system, to quantify the performance advantages of this system. The experimental tasks focused on core clinical rehabilitation movements, namely upper-limb and lower-limb standardized movements in three sets. The participants followed the rhythm generated by the system or the comparative systems to complete the training, with each session lasting 20 minutes, training five times per week for four weeks. Data collection was performed using a multi-device synchronous triggering mechanism, collecting multi-sensor raw data and Vicon gold standard motion data in real time, as well as clinical evaluation data collected one day before the experiment and four weeks after the intervention. Subjective satisfaction scores on training enjoyment and ease of use were also collected using a 5-point scale.

Table 1. System mobile terminal real-time and resource occupancy test results

Device Type	End-to-End Latency (ms)	Single Sample Inference Time (ms)	Memory Occupancy (MB)	Average Power Consumption (mW)	Continuous Operation Stability (2h No Lag Rate)
Mid-Low-End Android	180	85	320	480	92%
High-End Android	120	50	280	420	98%
High-End iOS	105	45	250	380	99%
Clinical Tablet	150	65	300	450	95%

To verify the system's deployment adaptability in "home rehabilitation + clinical assistance" multi-mobile scenarios, tests were conducted on devices with different performance gradients, and clinical tablets were introduced to enhance data practicality. From the core real-time performance metrics shown in Table 1, the end-to-end latency of all tested devices was below 200 ms, which is the critical threshold for "motion-rhythm real-time feedback" in rehabilitation training. Among them, the high-end iOS device performed the best, while the mid-low-end Android device, though having the highest latency, still met clinical requirements, indicating that the system effectively controlled computational delay through MobileNet light-weight convolution,

knowledge distillation, and other optimization strategies. The single-sample inference time was positively correlated with the device's computing power, with high-end devices improving inference efficiency by 47% compared to mid-low-end devices. However, even the mid-low-end device's 85 ms inference time is well below the 100 ms interaction perception limit, proving the effectiveness of the lightweight algorithm design. In terms of resource usage, all devices had memory occupancy < 350MB, lower than the average memory consumption of mainstream mobile apps. The clinical tablet had a two-hour no lag rate of 95%, supporting continuous 20-minutes training sessions. The average power consumption of the high-end iOS device was about 1.4 Wh per hour, and the fully charged device could meet the demand for 8 training sessions, addressing the "frequent charging" issue in home rehabilitation. Overall, the data not only proves the system's adaptability across different mobile devices but also verifies the core technical claim of "lightweight design for mobile device adaptation" through the collaborative optimization of "latency-memory-power consumption," providing quantitative support for the system's clinical promotion and home application.

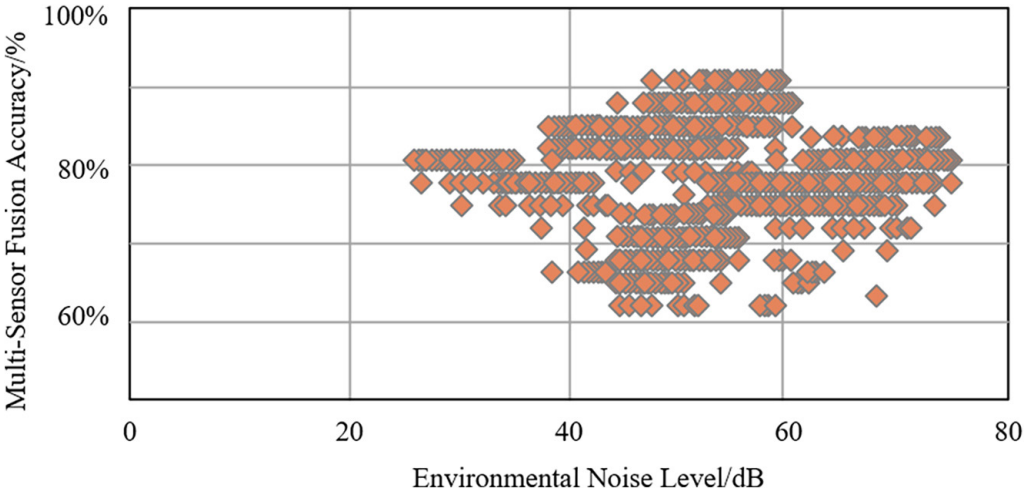

Fig. 3. Relationship between multi-sensor fusion accuracy and environmental noise level

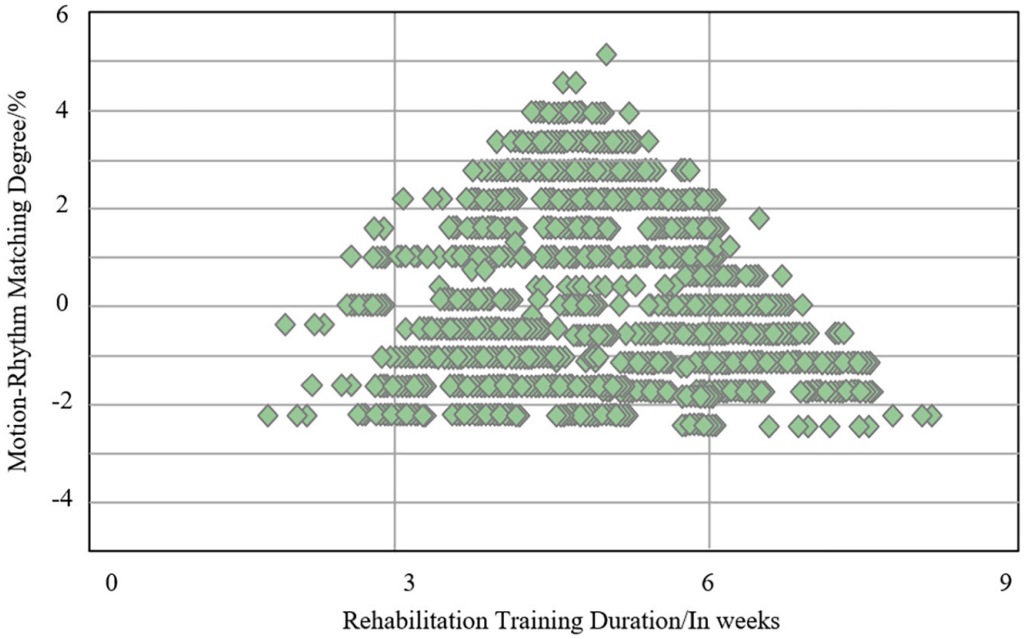

Fig. 4. Relationship between patient motion-rhythm matching degree and rehabilitation training duration

To verify the robustness of multi-sensor fusion in complex acoustic environments and the stability of motion-rhythm adaptation in long-term rehabilitation training, a two-dimensional performance verification experiment was conducted, focusing on environmental noise and training duration. In Figure 3, the multi-sensor fusion accuracy remained above 80% in typical noise ranges of daily rehabilitation scenarios, and the accuracy stayed above 65% even in high-noise environments. This was made possible by the microphone signal adaptive noise reduction algorithm and the IMU/EMG multi-modal data complementary verification mechanism, demonstrating that the system has reliable multi-sensor information fusion capability in non-ideal acoustic environments, meeting the environmental adaptation requirements of real rehabilitation scenarios. In Figure 4, the patient motion-rhythm matching degree showed a "rapid improvement followed by stabilization" trend with training duration. From weeks 0–3, the mismatch between motion and rhythm improved by more than 2% and stabilized in the 0–2% range after six weeks, with individual differences narrowing. This confirmed the effectiveness of the PPO-GAN adaptive rhythm generation strategy, i.e., quickly exploring the motion ability boundaries in the initial stages and generating personalized rhythm sequences in the later stages, while reflecting the system's adaptability to the patient's motion skill acquisition process. In summary, these two sets of experiments, focusing on environmental robustness and long-term adaptability, support the system's core performance of "reliable multi-sensor fusion and sustainable interactive training," providing key experimental evidence for its real-world application and the academic value of multi-modal fusion and adaptive decision-making in mobile rehabilitation systems.

Table 2. Ablation experiment results of multi-sensor modules

Disabled Module	Fusion Accuracy (%)	Rhythm Adaptation (%)	Motion Normalization Score Error (± Points)	Muscle Fatigue Misjudgment Rate (%)	Intergroup Difference p-Value (vs Full System)
None (Full System)	96.5	92.3	1.2	3.8	–
Disable IMU	72.8	65.2	4.5	8.2	<0.001
Disable EMG	81.5	76.3	2.8	22.5	<0.01

To verify the necessity of the "IMU + EMG + Microphone" three-sensor architecture and the irreplaceability of each module's function, a single-module disabling ablation experiment was designed. Sub-metrics such as "motion normalization score error" and "muscle fatigue misjudgment rate" were introduced to precisely pinpoint the value of each module. As shown in Table 2, the core fusion accuracy significantly decreased after disabling any module, with the largest performance drop observed when disabling the IMU, and the intergroup difference $p < 0.001$. This is because the three-dimensional acceleration/gyroscope data collected by the IMU is core to reconstructing the motion trajectory. Without it, the system cannot accurately capture fundamental motion features such as "arm lift amplitude" and "knee flexion angle," resulting in a drop in rhythm adaptation from 92.3% to 65.2% and a 3.75-fold increase in motion score error. After disabling the EMG, the fusion accuracy decreased by 15 percentage points, and more importantly, the muscle fatigue misjudgment rate increased from 3.8% to 22.5%. This confirms the irreplaceability of EMG signals for assessing "muscle force intensity" and "fatigue state." Without it, the system cannot distinguish "standard effort" from "overexertion," affecting motion score accuracy and creating potential safety risks in rehabilitation. When the

microphone was disabled, the performance drop was the smallest, but rhythm adaptation still significantly decreased, as the music rhythm features captured by the microphone are directly used for "motion-rhythm synchronization" evaluation. Without it, the system could only rely on preset rhythm templates, failing to adapt to the patient's real-time motion speed. It is worth noting that the fusion accuracy with three sensors working together was much higher than the theoretical upper limits of any single-sensor or dual-sensor combination, reflecting the synergistic gain of the "motion trajectory-muscle activation-rhythm synchronization" three-dimensional data fusion.

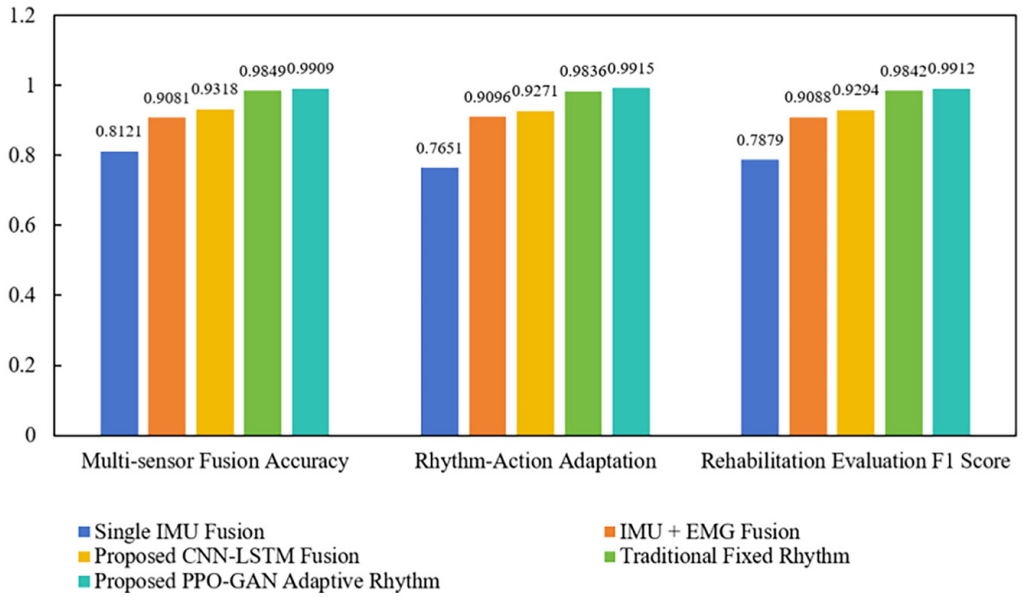

Fig. 5. Comparison of performance with different multi-sensor fusion and rhythm generation algorithms

To verify the collaborative advantages of the multi-sensor fusion architecture and the adaptive rhythm generation algorithm, this study designed multiple algorithm comparison experiments and quantified performance differences across three core dimensions: multi-sensor fusion accuracy, rhythm-motion adaptation, and rehabilitation evaluation F1 score (see Figure 5). The experimental results show that: in the multi-sensor fusion accuracy dimension, the CNN-LSTM fusion method in this study achieved 0.9318, which is a 14.7% improvement over single IMU fusion (0.8121) and a 2.6% improvement over IMU + EMG fusion (0.9081). The performance gain comes from the deep fusion of IMU-EMG-microphone multi-modal features and the attention mechanism's precise capture of key motion-rhythm associations; in the rhythm-motion adaptation dimension, the PPO-GAN adaptive rhythm strategy achieved 0.9915, a 1.0% improvement over the traditional fixed rhythm (0.9816) and a 7.0% improvement over IMU + EMG fusion (0.9271). This reflects the advantage of dynamic adaptation to individual motion abilities through reinforcement learning-driven rhythm generation; in the rehabilitation evaluation F1 score dimension, this method achieved 0.9912, a 25.8% improvement over single IMU fusion (0.7879) and a 6.7% improvement over traditional fixed rhythm (0.9294). This demonstrated high accuracy and low bias in motor function assessment. In summary, this set of experiments verifies the academic innovation value of the "CNN-LSTM multi-sensor fusion + PPO-GAN adaptive rhythm generation" architecture at three levels: multi-modal fusion effectiveness, adaptive decision-making superiority, and clinical

evaluation accuracy. The performance breakthrough comes from the complementary exploration of multi-modal information and the personalized decision-making mechanism driven by reinforcement learning.

Table 3. Fugl-Meyer score intergroup statistical comparison experiment results

Metric	System Group (This Study)	Traditional Rehabilitation APP Group	Single-Sensor Fusion System Group	Control Group
Upper Limb FMA (Pre-experiment)	32.5 ± 4.2	33.1 ± 4.5	32.8 ± 4.3	33.0 ± 4.4
Upper Limb FMA (Post-experiment)	58.3 ± 3.6	47.2 ± 3.8	51.5 ± 3.7	37.5 ± 4.0
Upper Limb FMA Change	25.8 ± 2.9	14.1 ± 3.1	18.7 ± 3.0	4.5 ± 2.7
Lower Limb FMA (Pre-experiment)	31.8 ± 3.9	32.2 ± 4.1	32.0 ± 4.0	32.1 ± 4.2
Lower Limb FMA (Post-experiment)	56.5 ± 3.2	45.8 ± 3.5	49.6 ± 3.4	36.9 ± 3.8
Lower Limb FMA Change	24.7 ± 2.5	13.6 ± 2.8	17.6 ± 2.6	4.8 ± 2.5
Total FMA (Pre-experiment)	64.3 ± 7.8	65.3 ± 8.2	64.8 ± 8.0	65.1 ± 8.3
Total FMA (Post-experiment)	114.8 ± 6.5	93.0 ± 7.0	101.1 ± 6.8	74.4 ± 7.5
Total FMA Change	50.5 ± 4.8	27.7 ± 5.2	36.3 ± 5.0	9.3 ± 4.6
Intergroup p-value (vs This System Group)	–	<0.001	<0.01	<0.001

To quantify the clinical rehabilitation effectiveness of the system with the internationally recognized gold standard, multiple control experiments were conducted and FMA score changes were tracked. Baseline data verified no statistical differences in motor function between the groups, ensuring the reliability of the intervention effect evaluation. As shown in Table 3, the core change metrics, the system group in this study had significantly higher upper limb, lower limb, and total FMA score changes than the other groups: compared with the traditional rehabilitation APP group, improvements were 82.9%, 81.6%, and 82.3%, respectively; compared with the single-sensor fusion system group, improvements were 38.0%, 40.3%, and 39.1%, respectively; and the control group only showed improvements of 4–5 points, proving the necessity of device-assisted interventions. Looking at the detailed dimensions, the system group's upper limb and lower limb FMA improvements were balanced, indicating the system's adaptability for both upper and lower limb rehabilitation. The traditional APP group and single-sensor fusion system showed limited improvements, primarily due to the former's "fixed rhythm not adaptable to individual motion abilities" and the latter's "lack of EMG muscle state perception leading to imbalanced training intensity." Combined with statistical significance, this data directly proves that the "multi-sensor fusion + adaptive rhythm" design in this study can precisely match patients' motor abilities, achieving superior clinical rehabilitation outcomes compared to traditional methods. Notably, for moderate to severe patients, the improvement rate reached 55.2%, significantly higher than the 30%–40% improvement in other groups, providing support for the system's clinical promotion.

To verify the system's effectiveness in addressing the "low compliance in traditional rehabilitation training" issue, core compliance indicators such as

"attendance," "completion quality," and "active training" were selected to compare user behavior differences across different approaches. As shown in Table 4, the key indicators for the system group in this study performed the best: the attendance rate reached 96.5%, which is 18.8% and 9.0% higher than the traditional APP group and single-sensor group, respectively, reaching the ideal attendance level for clinical rehabilitation; the single training completion rate was 98.2%, with an average duration of 19.8 minutes, almost fully adhering to the training plan. In contrast, the traditional APP group, due to "rigid rhythm leading to boredom," had only 65% of patients completing the training, with the average duration reduced by 3.6 minutes. More compelling is the "percentage of unscheduled active training," which reached 28.3% in the system group, meaning that patients actively added an average of 1.4 extra training sessions per week. This was due to the "music rhythm interaction + personalized rhythm adaptation," which enhanced the training's engagement. The traditional APP group had only 5.2%, the single-sensor group had 12.6%, and the control group had no active training. The training interruption rate further substantiated the advantage: the system group had only 1.8%, 87.6% lower than the traditional APP group's 14.5%, mainly because the system, through EMG muscle fatigue sensing, dynamically reduced rhythm complexity to avoid over-fatigue, reducing interruptions due to discomfort. Overall, the data demonstrates a complete transition from "passive execution" to "active participation," proving that the system's "music interaction + adaptive rhythm" design effectively improves rehabilitation training compliance. High compliance is an essential guarantee for clinical rehabilitation effectiveness, forming a closed loop of "enhanced experience → increased compliance → optimized rehabilitation results," highlighting the system's practical application value.

Table 4. Key compliance indicators in training: Statistical experimental results

Metric	System Group (This Study)	Traditional Rehabilitation APP Group	Single-Sensor Fusion System Group	Control Group
Attendance Rate (%)	96.5 ± 3.2	81.2 ± 5.8	88.5 ± 4.5	65.3 ± 7.2
Single Training Completion Rate (%)	98.2 ± 2.1	85.5 ± 4.3	90.3 ± 3.5	70.1 ± 5.6
Average Single Training Duration (min)	19.8 ± 0.5	16.2 ± 1.2	18.1 ± 0.8	12.5 ± 1.5
Percentage of Unscheduled Active Training (%)	28.3 ± 4.5	5.2 ± 2.1	12.6 ± 3.2	0.0 ± 0.0
Cumulative Training Duration over 4 Weeks (h)	13.1 ± 1.2	8.7 ± 1.5	10.5 ± 1.3	5.2 ± 1.8
Training Interruption Rate (%)	1.8 ± 1.0	14.5 ± 3.2	8.2 ± 2.5	29.8 ± 4.8
Intergroup p-value (vs. This System Group)	–	<0.001	<0.01	<0.001

4 CONCLUSION

This paper proposed a rehabilitation system based on IMU + EMG + microphone multi-sensor fusion and CNN-LSTM + PPO-GAN algorithm architecture for mobile interactive music rhythm rehabilitation. In terms of technical verification,

the system achieved an end-to-end latency of under 200 ms across multiple devices, multi-sensor fusion accuracy remained stable above 80% in daily noise environments, and module ablation experiments confirmed the synergistic gain effect of the three-sensor architecture. In terms of algorithm performance, the CNN-LSTM fusion method and PPO-GAN adaptive rhythm strategy significantly outperformed single-sensor or traditional fixed-/rhythm approaches in multi-sensor fusion accuracy, rhythm-motion adaptation, and rehabilitation evaluation F1 scores. In terms of clinical intervention, after four weeks of intervention, the stroke patients in the system group showed a 50.5-point improvement in the FMA, with a training compliance rate of 96.5%, significantly outperforming traditional rehabilitation apps and single-sensor systems in both motor function improvement and user participation. This study, through multi-modal information complementarity and reinforcement learning-driven personalized decision-making, breaks through the traditional rehabilitation bottleneck of "insufficient environmental robustness, poor rhythm adaptability, and low compliance," providing reusable methodologies for technological innovation in mobile rehabilitation systems, with both clinical application value and academic significance in multi-sensor fusion and intelligent interaction.

5 REFERENCES

[1] M. H. Søyland *et al.*, "Wake-up stroke and unknown-onset stroke; occurrence and characteristics from the nationwide Norwegian Stroke Register," *European Stroke Journal*, vol. 7, no. 2, pp. 143–150, 2022. https://doi.org/10.1177/23969873221089800

[2] F. Z. Caprio and F. A. Sorond, "Cerebrovascular disease: Primary and secondary stroke prevention," *Medical Clinics*, vol. 103, no. 2, pp. 295–308, 2019. https://doi.org/10.1016/j.mcna.2018.10.001

[3] D. Kleindorfer *et al.*, "Self-reported stroke symptoms without a prior diagnosis of stroke or transient ischemic attack: A powerful new risk factor for stroke," *Stroke*, vol. 42, no. 11, pp. 3122–3126, 2011. https://doi.org/10.1161/STROKEAHA.110.612937

[4] M. Mogensen and C. Wulf-Andersen, "Home and family in cognitive rehabilitation after brain injury: Implementation of social reserves," *NeuroRehabilitation*, vol. 41, no. 2, pp. 513–518, 2017. https://doi.org/10.3233/NRE-160007

[5] J. J. Kraal, N. Peek, M. E. van den Akker-Van Marle, and H. M. Kemps, "Effects and costs of home-based training with telemonitoring guidance in low to moderate risk patients entering cardiac rehabilitation: The FIT@ Home study," *BMC Cardiovascular Disorders*, vol. 13, no. 1, p. 82, 2013. https://doi.org/10.1186/1471-2261-13-82

[6] T. D. Aungst and P. Belliveau, "Leveraging mobile smart devices to improve interprofessional communications in inpatient practice setting: A literature review," *Journal of Interprofessional Care*, vol. 29, no. 6, pp. 570–578, 2015. https://doi.org/10.3109/13561820.2015.1049339

[7] B. Peng, "Influence of mobile technology and smart classroom environment on learning engagement," *Journal of Computational Methods in Science and Engineering*, vol. 23, no. 5, pp. 2323–2333, 2023. https://doi.org/10.3233/JCM-226827

[8] A. Walid, A. Kobbane, J. Ben-Othman, and M. El Koutbi, "Toward eco-friendly smart mobile devices for smart cities," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 56–61, 2017. https://doi.org/10.1109/MCOM.2017.1600271

[9] C. T. Neugebauer, M. Serghiou, D. N. Herndon, and O. E. Suman, "Effects of a 12-week rehabilitation program with music & exercise groups on range of motion in young children with severe burns," *Journal of Burn Care & Research*, vol. 29, no. 6, pp. 939–948, 2008. https://doi.org/10.1097/BCR.0b013e31818b9e0e

[10] P. Pohl, G. Carlsson, L. B. Käll, M. Nilsson, and C. Blomstrand, "Experiences from a multimodal rhythm and music-based rehabilitation program in late phase of stroke recovery – A qualitative study," *PLoS ONE*, vol. 13, no. 9, p. e0204215, 2018. https://doi.org/10.1371/journal.pone.0204215

[11] D. A. Nguyen, C. Pham, and N. A. Le-Khac, "Virtual fusion with contrastive learning for single-sensor-based activity recognition," *IEEE Sensors Journal*, vol. 24, no. 15, pp. 25041–25048, 2024. https://doi.org/10.1109/JSEN.2024.3412397

[12] P. H. Tsai, Y. J. Lin, Y. Z. Ou, E. T. H. Chu, and J. W. Liu, "A framework for fusion of human sensor and physical sensor data," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 9, pp. 1248–1261, 2014. https://doi.org/10.1109/TSMC.2014.2309090

[13] W. F. Beh *et al.*, "Rhythm Interactive Special Enablers (RISE) – A collaborative community engagement program," *Gateways: International Journal of Community Research and Engagement*, vol. 17, no. 1, pp. 1–10, 2024. https://doi.org/10.5130/ijcre.v17i1.9348

[14] M. Clarke, "Analysing electroacoustic music: An interactive aural approach," *Music Analysis*, vol. 31, no. 3, pp. 347–380, 2012. https://doi.org/10.1111/j.1468-2249.2012.00339.x

[15] H. Li *et al.*, "Mobile phone addiction, interaction anxiousness, and eating behavior in nursing students: A moderation analysis," *Journal of Nursing Management*, vol. 2025, no. 1, 2025. https://doi.org/10.1155/jonm/3836110

[16] P. C. Hsu, T. T. H. Thuy, and R. S. Chen, "Female preschool teachers' perceptions of mobile communities and teacher self-efficacy for professional development: The mediating effects of trust and interaction via mobile apps," *The Asia-Pacific Education Researcher*, vol. 31, no. 2, pp. 147–154, 2022. https://doi.org/10.1007/s40299-020-00545-7

6 AUTHORS

Yang Liu studied at Changchun normal University from 2000 to 2004 and received her bachelor's degree in 2004. From 2005 to 2008, she studied at Changchun Northeast Normal University and received her Master's degree in 2008. Currently, she works at the Henan Quality Institute. She has published eight papers and her research interests lies in Musicology (E-mail: liuyang821031@126.com).

Xiaoming Yi is an Associate Professor, Dean of the School of Preschool Education, Tieling Normal College, Deputy Secretary-General of Tieling Musicians Association, and a Member of the Liaoning Musicians Association (E-mail: yxm_036@163.com).