

Research on the Practice and Application Path of Bionic Robots in the Field of Piano Performance

Jiaen Gao

Hangzhou High School International Division, Hangzhou, 310021, China

Email: seangao2025@163.com

How to cite this paper: Gao, J. (2026). Research on the Practice and Application Path of Bionic Robots in the Field of Piano Performance. *Advances in Engineering Research: Possibilities and Challenges*, 4(3), 63–78. ISSN Print: 3079-5192, ISSN Online: 3079-5206.

<https://doi.org/10.63313/AERpc.9118>

Published: 2026-08-20

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

Traditional bionic piano-playing robots have long been confined to the fixed paradigm of "score reading - mechanical reproduction," suffering from three limitations: single perception method, scarce training data, and lack of creative and performing capabilities. This leads to the convergence of product functionalities, the flattening of user experience, and the difficulty of establishing effective market competitiveness. To address these issues, this paper has constructed an integrated intelligent music performance system of "perception - training - creation". At the perception layer, multimodal recognition combining acoustic recognition and visual recognition is introduced to achieve the mutual understanding and verification of score images and audio; at the data layer, passive mechanical exoskeletons are used to collect data such as joint angles and fingertip force of professional performers, and a force-emotion mapping dictionary and a spatiotemporal probability model for hand coordination are established; at the creation and performance layer, an artificial intelligence music creation module is designed, generating adapted scores that can be smoothly performed by the mechanical hand through style transfer and constraint optimization. Experimental verification shows that the accuracy of multi-modal score recognition has significantly improved compared to single visual recognition, the coordinated training data collected by the exoskeleton reduces the error in mechanical hand performance, and combined with the artificial intelligence composition module, it can stably perform multi-style adapted pieces. The core innovation of this paper lies in integrating the complete data loop from score recognition to performance, upgrading the bionic piano-playing robot from a fixed-score playback terminal to an intelligent performing entity with understanding, execution capabilities, and secondary creation. The research results provide a technical framework and a practical path for personalized design schemes and commercialization of intelligent music education.

Keywords

Piano-Playing Robot; Humanoid Robot; Force Control; Computer Vision; Intelligent Score Recognition; Deep Reinforcement Learning; Intelligent Composition

1. Introduction

1.1. Research Background

In the 21st century, humanoid robots have evolved from "mobile machines" to "embodied intelligent agents". In 2000, robots like Honda's ASIMO achieved bipedal walking, but their functions were limited. In the 2010s, the Atlas robot from Boston Dynamics enhanced its movement capabilities and could perform complex actions such as running and jumping. In the 2020s, AI large models drove the development of intelligent decision-making. New-generation robots like Tesla's Optimus possess natural interaction and autonomous learning capabilities and are gradually entering commercial applications. Domestic company Yushu Technology also publicly announced the research results of multi-robot cluster collaborative control technology. Based on the iterative optimization of hardware platforms, the deep integration of artificial intelligence technology has enabled humanoid robots to achieve a crucial technological leap from having excellent motion execution capabilities to also possessing intelligent perception and precise operational capabilities.

The piano has comprehensive technical expressiveness and is revered as the "King of Instruments," capable of simultaneously presenting melody, harmony, and rhythm elements, and can simulate the symphonic effects of an orchestra. However, mastering its playing skills requires overcoming a high technical threshold, and beginners need to undergo a long and systematic learning process. Piano art has deeply permeated into the daily life domain, with its repertoire covering almost all musical styles, providing a lifelong artistic exploration space for performers. Is such a complex instrument only manageable by humans? Currently, scientists are working on developing robots that can play the piano.

The integration of piano and robots is driven by various trends of the times, with the fundamental motivation being the advancement of service robots and artificial intelligence technology, which has driven the demand for emotional interaction robots. Piano playing as an emotional interaction method has become a key platform to test the performance of robots. The increasing demand for robots to perform complex tasks has made the high-precision and fast-paced scenarios of piano playing simulation an ideal training laboratory. Technically, from the automatic piano in the 19th century to the breakthroughs in motors, pneumatics, and AI algorithms in the 21st century, it has laid the foundation for robots to achieve expressive performances. This combination is not only an epitome of AI technology leaps, social demand drives, and core breakthroughs, but also marks the deep integration of intelligent systems in art and functionality.

1.2. Research Significance

In the field of music, this study aims to expand the boundaries of artistic expression through bionic piano-playing robot technology. The bionic piano-playing robot performance system is not constrained by physiological conditions and does not

have limitations on finger span or fatigue issues. It also has extremely high playing speed, enabling the realization of complex finger techniques and chord combinations that are difficult for humans to perform, and exploring the potential expressiveness of piano acoustics. Its "super physiological" performance characteristics provide composers with novel tone choices and creative concepts, and even facilitate the emergence of music works specifically composed for bionic piano-playing robots to perform. Additionally, this research has promoted the formation of new human-machine collaboration models, such as high-precision concerto or real-time interactive performance, thereby creating unexpected aesthetic sound effects. Therefore, this study is not intended to replace human performance, but rather extends the artistic limit through technological means, opening up new possibilities for music creation and performance.

From a commercial application perspective, bionic piano-playing robots, as a new type of interactive entertainment product, have significant market value. Robot bands can achieve stable and continuous performance of multiple pieces of music and have successfully been applied in large-scale events such as the Hangzhou Asian Games, confirming their feasibility in the exhibition economy and entertainment market. At the same time, these robots can be deployed in science and technology museums, public exhibition halls, etc., attracting audiences through performance and enhancing the experience, becoming a new commercial medium for the integration of technology and culture.

In terms of scientific and technological development, the research on bionic piano-playing robots has significant promoting effects. As the difficulty of playing increases, the requirements for precise hand control and multi-finger coordination continue to rise, prompting continuous iteration and upgrading of related technologies. Its core key technologies, such as high-precision motion control, multi-finger coordination algorithms, and real-time feedback systems, once mature, can be transferred to daily application scenarios, such as household service robots performing household tasks. Compared to the micrometer-level precision and millisecond-level coordination required for playing the piano, the technical requirements for daily tasks are lower, so the related technologies can be applied in a "downward" manner. This research not only aims to solve advanced machine intelligence problems in artistic performance but also significantly enhances the dexterous operation ability of the robots, providing key technological accumulation for the entry of service robots into the household environment.

1.3. Research Methods and Objectives

This study employs a systematic literature review approach, aiming to deeply explore the technical implementation paths and application prospects of bionic piano-playing robots. The specific research content covers the following three dimensions: Firstly, it conducts an in-depth analysis of the technical potential and

implementation mechanisms of artificial intelligence in music composition and secondary creation; Secondly, it assesses the feasibility of automatic score recognition technology and the key challenges it faces; Thirdly, it designs and discusses the high-precision control strategies for bionic mechanical hands in piano performance. At the same time, by reviewing the typical systems and core technologies of existing bionic piano-playing robots, it summarizes the current technical level achieved and the main bottlenecks. Finally, in combination with the latest advancements in the fields of artificial intelligence and humanoid robots, it makes forward-looking discussions on the future development trends and industrialization prospects of bionic piano-playing robots.

2. Literature Review

2.1. Development of Bionic Mechanical Hand Piano Playing Technology

The playing technology of bionic piano-playing robots in terms of hardware has been continuously updated and improved. Wei Xinyu proposed a force electromagnetic group system in the article "Piano Automatic Playing Mechanical Hand System and Its Control Algorithm Design" [8], using electromagnetic control and lever mechanics to precisely control the intensity of the playing. At the same time, a parallel coupling system algorithm was used to independently control the control of each key. However, in terms of bionic nature, there was no significant breakthrough and the posture design and coordination of bionic mechanical hands were lacking.

The complexity of the simulation-complexity composite still has many technical gaps. Zhao Yutong proposed a new idea in "Mechanical Exoskeleton Makes Pianists' Hands Move Like a Feather" [9], by combining the training of mechanical exoskeleton and brain scanning evidence, it was proved that the training changed the sensory cortex of the pianist, so as to better control the overall movement of the fingers. Although this technology still has not solved the core problem of bionic mechanical hands, its innovative design of exoskeleton assistance can be regarded as a transitional solution. It can also provide a data collection platform for big data human action posture capture and can to some extent reduce the complexity of the hands coordination algorithm.

2.2. Development of Music Notation Recognition Technology for Bionic Piano-Playing Robots

The development of music notation recognition technology for bionic piano-playing robots has gone through an evolution process from simple image recognition in the early stage to complex pattern recognition algorithms. Chen Xiaozhai et al. mentioned in their design of the Zhiyin Notation Recognition Mechanical Hand (知音识谱机械手的设计) (2018) that in the early stage, the research mainly relied on rule-based methods. Through preprocessing operations such as binarization, noise

removal, and edge detection on the music score images, basic elements such as notes and beats were extracted [7]. However, these methods have limited processing capabilities for complex music scores, especially when dealing with handwritten music scores or low-quality scanned images, where the recognition accuracy is difficult to meet practical requirements.

In the research process of music score recognition technology, how to convert paper five-line musical scores into computer-readable digital information has always been the core issue. Shi Wenjing (2023) proposed a complete music score image recognition system in his "Research and Development of Robot Notation Recognition and Piano Playing", stating that "multiple preprocessing operations were adopted to improve image quality, and the staff and notes were recognized using an end-to-end method. The recognition effects of two methods, LSTM and GRU, were compared, and the recognition results were reorganized and encoded. This system can achieve the entire process of music score recognition from input to output." [1] The experimental results show that LSTM (Long Short-Term Memory) outperforms GRU (Gated Recurrent Unit) in terms of training convergence speed and recognition accuracy, and can effectively reduce the error rate. Finally, the system reconstructs and encodes the recognized note information into MIDI format, providing a standardized digital music score basis for subsequent automatic playing tasks of bionic piano-playing robots.

2.3. Inefficiencies in Current Research

By comparing and analyzing the existing research, it can be observed that the current bionic piano-playing robots still have significant deficiencies in aspects such as score recognition accuracy, musicality in performance, and handling of complex pieces. Firstly, in terms of score recognition accuracy, although the application of deep learning algorithms has significantly improved the accuracy of music score recognition, problems such as overlapping notes and complex rhythms still constrain the performance of the system when dealing with multi-part music and polyphonic music [3][6]. For instance, the recognition method based on CRNN (Convolutional Recurrent Neural Network), although performing well in aligning the start and end points of notes, still requires improvement in the parsing ability of high-density note sequences [6].

Secondly, in terms of musicality in performance, the existing research mainly focuses on the technical implementation aspect, while the attention paid to artistic elements such as emotional expression and musical style is relatively insufficient [4]. Piano performance is not only the accurate reproduction of notes, but also requires a profound understanding of the style and cultural connotations of the work, which requires bionic piano-playing robots to have higher intelligence levels and musical literacy [4]. However, the performance systems of most bionic piano-playing robots still show a mechanical and monotonous nature in this aspect, making it difficult to

achieve the artistic appeal of human performers.

Finally, in terms of handling complex pieces, the existing research mainly conducts experiments on simple or moderately difficult pieces, and the support for complex scenarios such as large-scale symphonies or improvisational accompaniment is relatively limited [5]. For example, the problems exposed in piano improvisation teaching practice indicate that the existing systems have weak capabilities in responding to real-time arrangement and dynamic adjustment, making it difficult to meet the practical application requirements. These deficiencies provide an entry point for this research and point out the direction for the development of robot score recognition and piano performance technologies in the future.

3. Robot Music Notation Recognition Technology

3.1. Overview of Artificial Intelligence Image Music Notation Recognition Theory

Artificial Intelligence Image Music Notation Recognition (Optical Music Recognition, OMR) is an important research direction in music information retrieval. By combining computer vision and artificial intelligence technologies, it can convert paper music notation images into structured music data, complete intelligent transcription and digital processing of music notations, and is applicable to various scenarios such as music teaching, ancient music restoration, and music resource archiving.

The development of this technology is mainly divided into two stages. Traditional methods rely on machine vision algorithms such as template matching and edge detection to achieve noise reduction and basic note recognition. The operation is simple but has poor generalization and cannot adapt to handwritten, damaged, and complex layout music notations. The recognition accuracy is limited. After the rise of deep learning, end-to-end music notation recognition models have become the mainstream. Through convolutional networks to extract visual features of music notations and combining temporal networks to model symbol associations, it effectively overcomes problems such as multi-part superposition and image interference, significantly improving the recognition accuracy of various music notations.

Currently, this technology still has many shortcomings. The music notation symbols are complex, the layout is closely related, the handwriting styles of hand-written fonts vary greatly, and there are problems such as stains and blurriness in old music notations, which continuously affect the recognition effect.

3.2. Innovative Scheme of This Paper

This paper innovatively proposes that use Acoustic Recognition Assisting Visual Recognition. Traditional piano dynamic performance notations (strong, weak, crescendo, decrescendo, etc.) have ambiguous definitions, and the corresponding

performance boundaries and dynamic amplitudes are difficult to standardize and control. This notation system is designed based on human subjective interpretation expression. Music is an art that emphasizes emotional interpretation, and dynamic notations contain more than just the symbols themselves. The interpretation standards vary significantly depending on the work, and there is no unified and universal quantitative standard. At the same time, due to the different expressions of art, different people understand it differently, and the resulting effects vary significantly. Instead of having the machine understand human emotions, it is better to replicate the details of human emotional expression. Acoustic recognition has achieved this very well. During the music notation recognition process, both can be carried out simultaneously. The visual recognition is responsible for handling the specific pitch processing, while acoustic recognition captures real-time strength changes to improve the anthropomorphic emotional expression.

3.3. Technical Feasibility

The core technology that supports acoustic recognition assisting visual recognition lies in multimodal fusion. The system extracts acoustic features such as Mel-spectrograms and filter bank energy (FBank) from the audio, and visual features through Optical Music Recognition (OMR) from the images. Then, these two types of features are fused at different stages of the model. Subsequently, the audio and the musical score are matched, and the system can precisely locate the progress of the performance on the musical score by matching the audio segments and the musical score images. Finally, a unified representation space is established, allowing the three different forms of data—audio, musical score images, and symbolized musical scores—to be directly compared and retrieved in this space. For example, inputting an audio segment can retrieve the corresponding image in the musical score library. The multimodal system achieves a more powerful recognition effect than a single technology through its ability to fuse these two elements.

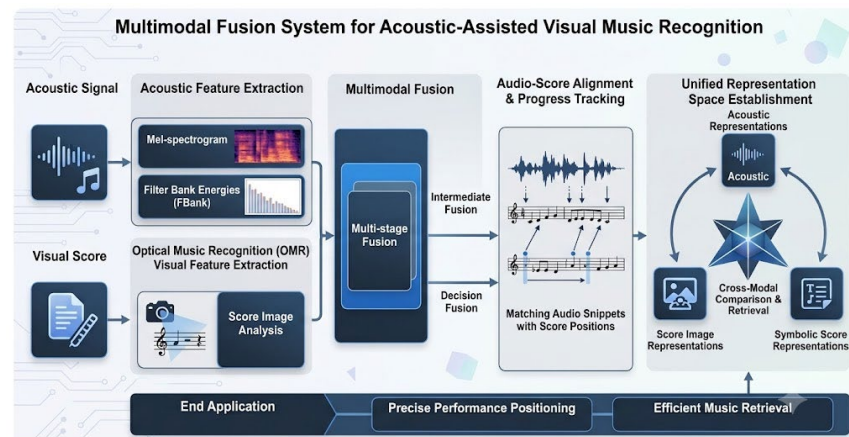


Figure 1: Flowchart of the Multimodal Fusion System for Enhanced Music Analysis

Figure 1. Flowchart of Acoustic Recognition Assisting Visual Recognition

4. Robot Piano Performance Technology

4.1. Core Technology Theory of Bionic Piano-playing Robots

The bionic piano-playing robot is an intelligent device that integrates bionic mechanical design and high-precision motion control. It can precisely replicate human playing movements and is widely used in intelligent performances, music education, and other fields. Its core technology is divided into two modules: bionic mechanical hand structure design and motion posture control.

4.2. Bionic Piano-playing Robot Bionic Mechanical Structure

The bionic mechanical hand is based on the bionic prototype of human playing limbs and adopts an industry-standard modular design. The hand is equipped with multi-degree-of-freedom bionic fingers and a micro servo drive structure, which can simulate finger flexion and button-pressing actions, adapt to the piano key distance and playing stroke, and combine with an elastic structure to avoid sound distortion. The arm adopts a multi-joint series structure, which can achieve translation, elevation adjustment, and meet the requirements of playing across eight octaves. The overall body adopts a lightweight and stable frame, optimizes the center of gravity layout, effectively suppresses high-speed playing vibration, and ensures the stability of the performance.

4.3. Motion and Posture Control Technology of Bionic Piano-playing Robots

Motion and posture control are the key to precise performance. The system relies on MIDI (Musical Instrument Digital Interface) sheet music data to complete real-time motion trajectory planning, intelligently allocates key sequences, and adapts to complex musical rhythms. Through real-time collection of posture and stroke data by multiple sensors, it dynamically corrects mechanical errors, ensuring precise key positions. At the same time, it uses a force control closed-loop algorithm to regulate the falling speed and torque of the keys, precisely distinguishes the strength of playing, and combines trajectory optimization algorithms to suppress playing tremors and lags, achieving stable, smooth, and in harmony with the emotional content of the music high-precision performance.

4.4. Specific technical support and feasibility plan for the posture control training of bionic mechanical hands using exoskeleton mechanical hands

This paper will adopt the exoskeleton mechanical hand assistance information collection method mentioned earlier. Experienced pianists often rely on intuition and experience when assigning finger positions, and it is difficult for them to express it completely in words and rules. Therefore, the exoskeleton motion capture method is adopted. Through a large number of passive movements for playing, the

exoskeleton mechanical hand can record the finger position design required when handling different spans, and capture the fine adjustments of fingers when handling complex and difficult notes. In terms of hand coordination, the two exoskeleton mechanical hands can design observation and positioning points at each joint and landing position, and then be comprehensively captured by external cameras in a three-dimensional space to form a motion trajectory analysis, which greatly reduces the training cost and algorithm difficulty of traditional bionic mechanical hands.

4.4.1. Specific implementation technical plan of exoskeleton mechanical hand

This technology differs from the traditional scheme in that the exoskeleton does not exert force but is actively pulled by the human hand, and at the same time collects the light and heavy force, sustains time, left and right hand system coordination, and the optimal solution for complex finger changes. First, angle and motion trajectory sensors will be set on the joints of the exoskeleton mechanical hand. By integrating finger joint angle sensors on the exoskeleton mechanical hand, the bending angle, motion trajectory, and angular velocity of each joint during playing can be accurately captured, and at the fingertips, a piezoresistive sensor array can be integrated to capture the slight changes in pressure.



Figure 2. Flowchart of exoskeleton mechanical hand

4.4.2. Data Collection and Conversion of Bionic Piano-Playing Robot Exoskeleton Mechanical Hand

Specifically, the collected raw data needs to be analyzed and converted into signals that can be used by the bionic mechanical hand. In terms of force collection, the fingertip pressure array data is aggregated into the total force value for a single key press (the integral of pressure at each pixel point of the sensor), and is divided into five levels of piano strength and softness classification: pp (extremely soft, 0–20% of the bionic piano performance robot), p (weak, 20–40%), mf (moderate strong, 40–60%), f (strong, 60–80%), ff (extremely strong, 80–100%). At the same time, for each collected force data, a label related to the musical theme and emotion should also be attached. Using structural recognition algorithms, the performance recording is automatically segmented into sections, and the section type (main

verse/verse, etc.) is marked. Thus, all the force data within a section can be uniformly labeled with a structural tag. By collecting a large amount of annotated emotional and thematic classical piano pieces for performance data, the statistical values of force for each quarter or section (average value, variance, change rate, etc.) can be extracted. A classifier (such as a simple neural network) is trained, with the input being the force statistical features and the output being the probability distribution of emotional themes. Further, an end-to-end force generation model is trained, with the input being style emotion and theme information, and the output being the predicted force value for each note.

4.4.3. Data Processing of Exoskeleton Mechanical Hand (Coordination Constraint Modeling of Both Hands)

The core solution of this technology is to convert the physical space information collected by the exoskeleton mechanical hand into computable mathematical models: three-dimensional coordinate mapping. Each finger and each key are placed in a coordinate system where the x-axis represents the position of the 88 keys, the y-axis represents the extension length of the finger from the palm to the tip, and the z-axis represents the height of the finger from the key. Thus, at any time t , all the finger joints of the left or right hand can be represented by a set of three-dimensional coordinates. Each key is also defined with a three-dimensional coordinate (the center point of the finger that is frequently pressed on the key surface). With this coordinate system, we can calculate the Euclidean distance (the straight-line distance in three-dimensional space) between the finger and the key at any time, as well as the relative spatial relationship between different fingers.

Next, for different performance modes (synchronous, alternating, melody-accompaniment, etc.), separate spatio-temporal probability distribution models are established for them. For example, analyzing the trajectory data of both hands when the performer alternates rapid scales, and statistics on which moments the distance between the hands is the closest, and what the distribution of the closest distance is. These statistics are converted into a Markov state transition matrix to describe the probability patterns of the hands alternating in different regions.

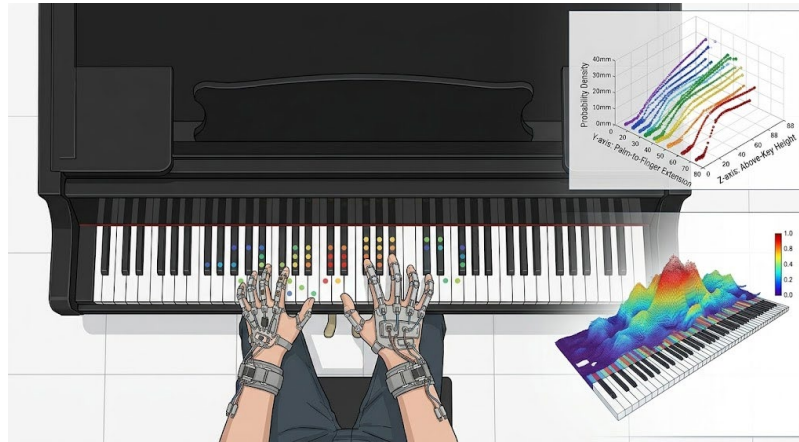


Figure 3. Schematic diagram of data processing and conversion for exoskeleton mechanical hand

With the above probability model in place, what the system needs to do during real-time performance is: based on the current position of the music score and the hand states from the previous moment, it needs to infer online which mode should be adopted currently and dynamically adjust the probability distribution. The system will output a mode probability distribution to guide the subsequent path planning. For example: there is a 70% probability that the current musical phrase is in the synchronous mode and a 30% probability is in the alternating mode. Then the system will weight-average the optimal paths of these two modes. At the same time, it adds constraints such as collision constraints (determining a collision to occur with a distance lower than a certain threshold) and mechanical limit constraints (limitations on the torque and speed of the mechanical hand motor), and converts the penalty function into the posture simulation function. If the trajectory violates the constraints, a huge penalty value will be added, thereby making the optimization algorithm tend to choose a trajectory without collisions.

5. The Technological Potential and Implementation Mechanism of Artificial Intelligence Music Creation

5.1. Research Positioning and Value of the Arrangement Creation Module

The previous text has respectively discussed the implementation paths of AI music notation recognition technology and the bionic robot piano playing system: The former uses Optical Music Recognition (OMR) and audio signal processing to convert paper sheet music or audio samples into structured digital representations through computer vision and artificial intelligence technologies; the latter uses exoskeleton motion capture combined training to collect data. Together, they have constructed a complete path from sheet music to performance. However, in the hierarchical architecture of the intelligent music system, music notation and performance functions still belong to the basic level of "perception - execution"; the

expansion of their higher-level capabilities must point to the "creativity" dimension. Through the combination of artificial intelligence music creation and traditional bionic piano playing robot performance technology, it has crossed the basic logic of imitation-based performance of bionic pianos, and provided a new paradigm from "virtual generation" to "physical performance" for digital music production, offering an operational experimental paradigm for re-examining the essential attributes of artificial intelligence creativity and its artistic creation measurement standards.

5.2. Overview and Existing Shortcomings of Artificial Intelligence Music Creation Technology Development

Artificial intelligence music creation has undergone a paradigm evolution from probabilistic generation to deep learning. In the 1950s, some music generation systems based on rules and probabilities emerged. In 1957, Hiller and Isaacson used Markov chains to generate the string quartet "Iliac Suite", opening the way for computer music generation. Since then, although probability methods such as Markov models can generate basic melodies through the connections and dependencies between notes, they are unable to capture the long-range structure of music, lack emotions and themes, and the melodies are relatively thin. Later, a music random combination method emerged, dividing music into small segments and recombining them according to features. In 1994, Mozer introduced neural networks into music generation, and deep learning has since become the core driving force of this field. Long short-term memory networks, by capturing music features, perform significantly better than traditional recurrent neural networks in generating coherent melodies, while generative adversarial networks, restricted Boltzmann machines, and other models have further expanded the boundaries of multi-track music generation.

However, current deep learning models generally rely on large-scale labeled data, and their generalization ability is insufficient when faced with limited samples of specific music styles. Wang Siwen proposed a new paradigm of meta-learning in the application of artificial intelligence composition in the article "The Application of Meta-Learning in Artificial Intelligence Composition" (2024), enabling the model to quickly extract style features from a small number of samples and adapt to new tasks [10]. The algorithm studies based on the meta-learning idea show that the MAML algorithm, compared with traditional algorithms, has core advantages of being architecture-independent and significantly improving generalization ability, and can enable the model to quickly converge and iterate using a small number of samples when facing new music styles, effectively solving the problem of difficulty in style transfer for traditional models under the condition of few samples, and providing a new theoretical path for more personalized music generation. The paper also mentions the shortcomings of this technology, that the current music generation is mainly limited to single-note sequences and fails to effectively handle

more complex chord generation, and the current research does not involve this complex theory.

5.3. Innovative Artificial Intelligence Music Creation Solution for Adapted Bionic Piano-Playing Robots

Traditional artificial intelligence music creation is based on recompilation of existing music materials, which is complex and difficult to precisely control the direction of the music. This paper proposes a secondary arrangement solution. Through artificial intelligence music composition, based on the original music piece obtained by intelligent music reading, it is adapted and can realize the change of style according to the user's requirements. At the same time, combined with the parameters such as the intensity, sustain time, and pedal depth generated by the robot hand during the performance, as the conditional constraints of the secondary arrangement model. Thus, it is not limited to static music score generation, but can dynamically adjust the harmony density, part balance, and texture level based on the real physical feedback of the actual performance.

5.4. Technical Implementation Path of the Innovative Solution

In terms of mechanism, we introduce a dual-channel encoding and decoding architecture: The encoder integrates the note sequence obtained from music reading and the performance dynamics features provided by the robot hand, and extracts the joint representation of "score structure" and "performance expression" through the cross-modal attention mechanism; The decoder adopts a conditional diffusion model, while retaining the original melody framework, it can controllably redraw the chords, rhythm patterns, and orchestral colors according to the specified style parameters (such as jazzization, orchestration) by the user. What is particularly innovative is that the composition result can drive the robot hand to perform a visual and auditory replay again. The system iteratively optimizes the composition parameters by comparing the error between the expected and measured audio, forming a "music reading → performance → analysis → adaptation → re-performance" perception-action-creation loop. This mechanism not only breaks the traditional linear process of composition and performance, but also enables artificial intelligence to acquire the subtle dimensions of expressiveness from each physical keystroke, allowing artificial intelligence music creation to rise from pure data calculation to an intelligent practice with physical constraints and aesthetic feedback, providing unprecedented interaction depth and artistic possibilities for digital music creation.

5.5. Industrialization and Economic Benefit Analysis of the Artificial Intelligence Music Creation Module

Supported by the "perception - training - performance" full technical loop, this

artificial intelligence music creation system demonstrates broad application prospects in multiple vertical scenarios. In the fields of short-video and streaming media content production, the system can quickly adapt the melody fragments uploaded by users into multiple versions of accompaniment music that suit different needs and themes, significantly improving the efficiency of content creators in providing background music; in the fields of film and game music, the system parses the reference music features and automatically generates multiple style adaptation plans, significantly reducing the cost of custom music composition for small-scale projects, providing rapid iteration assistance for development teams; in music education and teaching assistance, the system can systematically build a "demonstration performance - score analysis - intelligent adaptation - mechanical hand re-demonstration" closed loop, automatically generating simplified or technical practice versions of the pieces and physically performing them, effectively enriching teaching methods.

From a commercial perspective, the global artificial intelligence music generation market is expanding at a compound annual growth rate of nearly 30% and is rapidly growing. A certain short-video platform has adopted an artificial intelligence generation solution for over 70% of its background music, and the number of users of domestic artificial intelligence music creation platforms has exceeded 5 million. The diversified profit paths such as subscription, API services, and enterprise customization are becoming increasingly mature. This system, with its core feature of combining "virtual generation" and "physical performance," is expected to occupy a differentiated ecological position in the tens-of-billion-dollar market and promote the evolution of artificial intelligence music from an auxiliary tool to a creative infrastructure.

5.6. Differentiated Enabling Value of the Artificial Intelligence Music Creation Module for Bionic Piano-Playing Robot

Currently, the commercially available bionic piano-playing robots are generally limited to the fixed paradigm of music score reading and mechanical reproduction, with highly similar product functions and difficult to meet users' demands for personalized music experiences, nor can they form effective brand differentiation. The embedding of the artificial intelligence music creation module fundamentally breaks this limitation: based on the traditional mode of reading the score and outputting, through style migration and diverse tone color adaptation, the same piece of music can be instantly generated into jazz, pop, minimalist, etc., multiple performance versions, and according to the performance scenario (teaching demonstration, entertainment accompaniment, background music), the arrangement strategy is dynamically adjusted, significantly expanding the usage dimensions of the product. More importantly, the composition module is trained based on the data collected by the exoskeleton mechanical hand, so the feasibility of

finger positions, the comfortable range of force, and the constraints on hand coordination are taken as generating conditions, ensuring that the adapted results can be smoothly executed by the mechanical hand, eliminating most of the compatibility problems of artificial intelligence music creation systems and secondary development platforms in the market. This "understanding - re-creation - execution" integrated capability enables the bionic piano-playing robot to upgrade from a fixed music playback terminal to an interactive and customizable music companion, building a distinct technical barrier and brand premium space in the highly homogeneous market.

6. Conclusion and Outlook

This paper systematically reviews the core technological progress of bionic piano-playing robots and proposes a comprehensive technical framework that integrates auditory-assisted visual perception, bionic training deep learning, and ecological adaptation creation generation. The aim is to achieve a full process integration from the extraction of five-line staff information to high-precision mechanical performance and secondary creation.

By reviewing and analyzing existing literature, this paper points out the bottlenecks faced by each independent field technology in practical applications. In the field of visual music score recognition, due to the limitations of traditional optical character recognition (OCR) in complex chord discrimination and special symbol recognition, this paper proposes a multimodal fusion scheme of acoustic recognition assisting visual recognition, effectively improving the accuracy and robustness of music score parsing. In the field of bionic mechanical hand control, addressing the problems of stiff touch keys in traditional rigid mechanical mechanisms, slow multi-finger collaborative response, and the difficulty in collecting posture training data, this paper designs an acoustic recognition assisting visual recognition to collect finger techniques and emotional expression data and predict and analyze the characteristics of strength and weakness, significantly improving the dynamic response and emotional expression of the robot during performance. In the field of intelligent music creation, addressing the pain point that the music generated by existing algorithms often lacks consistency in long passages, this paper introduces a dual-channel encoder-decoder architecture that changes the timbre according to user needs and adapts to the working range of the mechanical hand, enhancing the performance of the algorithm in the extension of musical motives and overall arrangement.

This paper demonstrates that these three modules are not simply a simple functional superposition, but have formed an efficient collaborative "perception-decision-execution" closed-loop system. The establishment of this system breaks through the functional limitation of traditional playing robots that can only perform "mechanical MIDI file playback", giving the system higher-order

artistic expression ability and environmental adaptability, providing a practical reference value theoretical basis and technical path for the research and development of music education-assisted robots and highly bionic interactive systems.

Although this paper has made certain progress in system framework integration and key technology improvement, there are still many challenges. For example, in the mechanical response delay under extremely fast note sequences, and in the oscillation and balance of fitting bionic human form in the entire body movements, there is still room for further improvement. Future research will focus on more lightweight and high-bandwidth mechanical driving materials, with the aim of achieving more natural, delicate, and emotionally charged bionic robot performance.

References

- [1] Shi Wenjing. Research and Development on Robot Notation Recognition and Piano Performance [D]. Guangdong: Guangdong University of Technology, 2023.
- [2] Sun Shan'an. Design of Piano Automatic Performance Robot under Infrared Wireless Transmission [J]. Wireless Internet Technology, 2023, 20(23): 1-5.
- [3] Hu Limin; Gui Hao. Research on the Application of Artificial Intelligence in Popular Piano Education [J]. Art Evaluation, 2021, (7): 101-103.
- [4] Zheng Wei. Training of Piano Performance Skills and Cultivation of Artistic Literacy - Review of "Dancing on the Piano Keys: Piano Performance and Teaching Practice" [J]. Chinese Journal of Education, 2023, (8): 10043-10043.
- [5] Liu Ming. Practical Thoughts on Piano Improvisational Accompaniment Teaching in Music Education [J]. Hebei Picture Magazine, 2024, (16): 120-122.
- [6] Hu Limin; Gui Hao; Du Jijun; Tang Jianxiong; Chen Kaiyi. Research on the Method of Piano Music Transcription at the Note Level Based on Convolutional Neural Network [J]. Computer Application and Software, 2022, 39(5): 179-185.
- [7] Chen Xiaozhao; Sun Jingwei; Zou Shuichao, et al. Design of Piano Notation Recognition Mechanical Hand [J]. Laboratory Research and Exploration, 2018, 37(11): 81-85. DOI: 10.3969/j.issn.1006-7167.2018.11.021.
- [8] Wei Xinyu. Mechanical Hand System for Automatic Piano Performance and Its Control Algorithm Design [J]. Journal of Jingdezhen University, 2023, 38(6): 32-37.
- [9] Zhao Yutong. Mechanical Exoskeleton Makes Piano Players' Hands Move Swiftly [J]. China Science and Technology Daily, 2025-01-20(002).
- [10] Wang Siwen. Application of Meta-Learning in Artificial Intelligence Composition [D]. Shanghai: Shanghai Applied Technology University, 2024