2024 IEEE International Conference on
Advanced Robotics and Its Social Impacts (ARSO)
Hong Kong, China, May 20-22, 2024

2024 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO) | 979-8-3503-4463-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ARSO60199.2024.10557797

Humanoid Pianist: Dexterous Piano Playing with Synergy-based Hand Representation*

W. Jason Li[1,2], Zhuo Li[1], Junjia Liu[1], Zhipeng Dong[1], Zheng Sun[1] and Fei Chen[1], Senior Member, IEEE

***Abstract*— Playing the piano with dexterous hands has long presented unique challenges for robotics research due to the complex dexterity required. Traditional approaches utilizing individual control of each finger and joint in a robot hand make achieving fluent musical performances difficult and introduce significant control complexity due to the high degree-of-freedom nature of the dexterous hand workspace. This paper introduces a system that explores dexterous piano playing more naturally through a synergistic approach. The system employs a dual-armed robot with hands designed around two main synergies governing coordinated finger movements, which allows defining simple poses to represent both single-note and double-note playing. A mapping connects specific musical notes to the corresponding hand poses and precise timing required for a realistic piano rendition. A robot operating system is utilized to allow complex sequences and rhythms to be concisely specified. Results show that a synergistic approach is feasible and allows convenience with considerable accuracy and repeatability.**

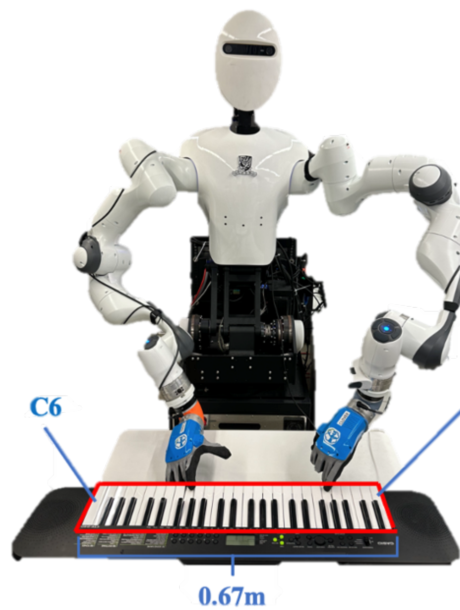

Fig. 1. Platform overview of the humanoid pianist: The Collaborative dUal-arm Robot manIpulator (CURI) robot with Franka dual arms and qb SoftHand2 is employed to play the electronic piano. The system setup is shown with playable range highlighted in red.

I. INTRODUCTION

Music plays a significant role in the lives of many individuals, serving as an essential and enjoyable activity that is integrated into their daily routines [1]. With its captivating melodies and harmonies, it has the capacity to evoke emotions, foster cultural connections, and serve as a conduit for personal expression. The emergence of robotic musicians represents a remarkable fusion of technology and artistry. Nowadays, many robots already possess the ability to replicate the motions and dexterity required to play musical instruments with proficiency and precision.

By combining advances in robotics technology with the nuanced requirements of musical expression, these studies strive to bridge the gap between human musicians and their mechanical counterparts. Advancing robotic systems capable of performing piano music with remarkable precision, dexterity, and expressiveness can bridge the gap between human musicianship and robotic performance, pushing the boundaries of what machines can achieve in the realm of musical expression. This broader picture encompasses a rich history of exploration and innovation, showcasing the ongoing progress in developing robotic systems capable of captivating piano performances.

Synergy-based robot hands have gained significant attention in the field of robotics due to their ability to mimic coordination and dexterity of the human hand. A synergy-based robot hand concept involves the coordination of multiple joints to achieve specific hand gestures in robotic design and control [2].

These robotic hands are designed and controlled in a way that allows multiple joints to work together seamlessly, enabling the execution of specific hand gestures with precision and efficiency. To achieve this level of coordination, researchers have sought to understand the underlying principles and patterns of human hand coordination. By studying the neuroscientific foundations of hand synergies, they have gained insights into how these principles can be translated into robotic hand designs and control mechanisms.

In this study, we explore the feasibility of employing a humanoid robot with synergy-based robot hand for piano playing. Unlike traditional robot hands that necessitate individual control over each finger and joint, the synergy-based robot hand utilizes only a limited number of commands

*This work was supported in part by the Research Grants Council of the Hong Kong SAR under Grant 24209021, 14222722, 14211723, and C7100-22GF and also in part by InnoHK of the Government of Hong Kong via the Hong Kong Centre for Logistics Robotics. (Corresponding author: Fei Chen.)

[1]W. Jason Li, Zhuo Li, Junjia Liu, Zhipeng Dong, Zheng Sun and Fei Chen are with the Department of Mechanical and Automation Engineering, T-Stone Robotics Institute, The Chinese University of Hong Kong, Hong Kong (e-mail: zli@mae.cuhk.edu.hk; jjliu@mae.cuhk.edu.hk; zhipeng-dongneu@gmail.com; zhengsun@link.cuhk.edu.hk; f.chen@ieee.org).

[2]W. Jason Li is also with the Faculty of Engineering, The University of Hong Kong, Hong Kong (e-mail: liwaynejason.1@gmail.com).

within the geometric synergy coordinate range to effectively reach the intended hand poses. Consequently, considering the distinct flow of robot hand movement compared to that of the human hand, exploring the available motions within the constraints of this hardware presents an intriguing research avenue. Hence, by undertaking this investigation, we aim to further our understanding of synergy-based dexterous manipulation in practical applications. We summarize our contributions as follows:

- Our work showcases the remarkable capability of a humanoid robot to play the piano, spanning 4 octaves of keys, achieved through the utilization of a synergy-based robot hand.
- We mapped the specific poses available for piano playing within the geometric synergy coordinate range, allowing us to precisely define the range of poses that the robot hand can execute to accurately replicate motions required for piano playing.

II. Related Work

During our preliminary research, we conducted a comprehensive examination of existing work in two key domains: robotic piano playing and synergy-based dexterous manipulation.

A. Robotic Piano Playing

There has been a long-standing, i.e. since year 2004 [3], dedication to the field of music instrument playing using robotics. Numerous studies have focused on designing specialized hardware and customized controllers tailored specifically for playing songs. These approaches involve the use of pre-programmed commands, audio feedback, and self-calibration (e.g., see [1], [3], [4], [5], [6]). This rich history demonstrates the ongoing efforts to create robotic systems capable of performing piano music.

The Mini-Humanoid Pianist [1] exhibits the ability to play scales or simple melodies imported from a Musical Instrument Digital Interface file. It possesses the capability to detect and correct inaccurately pressed keys. With an impressive overall accuracy rate of 99.6% and a maximum speed of 200 beats per minute, the Mini-Humanoid Pianist demonstrates remarkable proficiency in piano playing. However, it is worth noting that its robot hand is limited to a single finger with a hooked shape, deviating from the multi-fingered structure of a human hand. In a similar vein, [5] presents a robot that responds to sound feedback. This robot incorporates a custom end-effector equipped with a 3D printed finger, featuring a flat origin and a rounded finish. The design allows for more precise interaction with the keyboard, enhancing the robot's playing capabilities. Alternatively, the Keyboard-playing robot described in [3] employs linear motors and pneumatic cylinders on two human-like 5-fingered hands. The control mechanisms are hard-coded, providing the robot with the necessary motor control to manipulate the keys on the keyboard.

To sum up, the aforementioned related works highlight different approaches and advancements in robotics for keyboard playing. These studies collectively contribute to the ongoing progress in developing robotic systems capable of performing intricate tasks in music instrument playing.

B. Synergy-based Dexterous Manipulation

In [2], the authors present a comprehensive overview of the neuroscientific foundations of hand synergies and delve into the understanding of how these synergies can be applied to robotic hand designs. The article highlights the application of hand synergies in robotics by presenting examples, including the Pisa/IIT SoftHand. These examples demonstrate how the principles of hand synergies can be implemented in robotic systems to enhance their dexterity and versatility. While in a related work [7], the authors focus specifically on tasks involving actual hand movements and different types of grasping. They explore how the concept of synergy can be applied to maintain force equilibrium through the fingertips during grasping actions. This research sheds light on the practical applications of hand synergies in robotic systems, providing valuable insights for the development of advanced robotic grasping capabilities.

Drawing inspiration from how the human Central Nervous System manipulates their hand [2], robotics researchers have attempted to conduct research on the development of synergy-based robotics systems. The articles relate to hand synergies in robotics are insightful resources for examining the feasibility of the application of synergy-based robot hand in the field of humanoid piano playing.

III. Robotic System

A. Robot Platform

In this work, the Collaborative dUal-arm Robot manIpulator (CURI) robot platform is utilized, which consists of a total of 22 Degrees of Freedom (DoF), excluding the hands. The Franka dual arms are employed, each offering a maximum reach of 855 mm and equipped with 7 DoF, featuring torque sensing at each joint [8]. The hands employed are the qb SoftHand2, which possess 19 anthropomorphic DoF and incorporate two synergies [9]. The first synergy facilitates a continuous array of poses, spanning from extending the thumb and index finger to extending the little finger. On the other hand, the second synergy allows for a continuous range of poses, transitioning from a fully opened hand to a closed one. These synergies work harmoniously to enable the hand to execute a range of poses, which is detailed in section IV-B.

B. Robot Hands

During our initial analysis, we found that it is indeed feasible to utilize the qb SoftHand2. This is due to the fact that the average width of white keys on a typical keyboard is approximately 23.5 mm, while the black keys have an average width of 13.7 mm [11]. Considering the width of the robot hand fingers being approximately 20 mm, we have confidence that our robot hands can accurately play

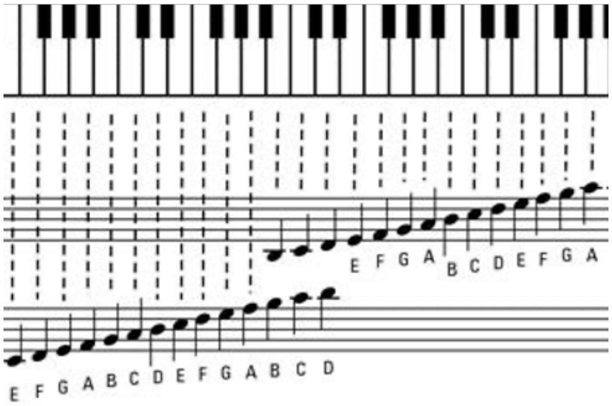

Fig. 2. The correspondence of piano keys to piano notes on the treble clef (notes for right hand playing) and bass clef (notes for left hand playing) [10].

the desired notes without inadvertently pressing neighboring keys.

However, one of the key limitations of our application is the restricted number of poses that the two synergies of the robot hand can generate, specifically for piano playing. While the hand is primarily designed for grasping tasks, it lacks independent control over the thumb and index finger, i.e., it is impossible to move the thumb while simultaneously preventing the index finger from moving. This same limitation applies to the middle finger, ring finger, and little finger, where the movements of these fingers are interdependent.

The limitation is not significant when playing single notes, where only one piano key needs to be pressed at a time. However, the disadvantage becomes apparent when it comes to playing double notes that span an interval larger than a 5th, i.e., in most cases four white keys apart.

The maximum playable width achievable with a single hand will be restricted. For instance, it is not feasible to press the keyboard using both the thumb and middle finger simultaneously. In such cases, either the other hand needs to assist, or the note cannot be played accurately.

C. Communication Interface

The Franka Control Interface (FCI) serves as a crucial tool in facilitating a fast and direct low-level bidirectional connection between the Franka Arm and qb SoftHand2 [12]. By leveraging the capabilities of the FCI, real-time control values on the Franka dual arms can be achieved at a speed of 1 kHz. This remarkable speed empowers seamless movement of the arm, enabling precise positioning according to desired coordinates.

In parallel, the qb SoftHand2 Research Graphical User Interface (GUI) assumes a pivotal role in controlling the inputs of two synergies [14]. Through the utilization of the advanced control tab within the GUI, we gain the ability to exert precise control over individual motors, thereby enabling intricate manipulation tasks [14]. Poses can be reached by simply by adjusting the two synergies in the geometric coordinate range. This feature-rich interface empowers researchers to explore and fine-tune the capabilities of the qb SoftHand2, expanding the realm of possibilities in dexterous manipulation experiments.

TABLE I
THE CORRESPONDENCE OF THE NOTES AND RESTS TO THE RELATIVE DURATION [13]

Note	Rest	Relative Duration
♩	𝄽	1
𝅗𝅥	𝄼	2

Furthermore, the integration of the Robot Operating System (ROS) plays a vital role in streamlining the flow of information between sensors and actuators [15]. By swiftly connecting the data received from sensors, ROS facilitates the execution of commands issued by control systems, enabling a seamless and efficient interaction between perception and action. This seamless integration of ROS into the robotic setup ensures a cohesive and synchronized workflow, enhancing the overall performance and responsiveness of the system.

IV. METHODOLOGY

A. Overview of Pitches and Rhythms

A comprehensive understanding of the music score is crucial for accurately interpreting and enabling the robot to perform the intended music.

In a music score, the notes represented on the five-line staffs indicate the pitch (frequency) of each note. The higher a note is positioned on the staff, the higher its pitch. Hence, as shown in Figure 2, the notes can be directly mapped from the music score to the electronic piano. Subsequently, the position of the robot's arm and hand can be adjusted to reach the desired location.

Reading the music score vertically provides information about the pitch of the notes, while reading horizontally reveals the rhythm (duration) of the notes. For simplicity, the following table displays only two of the most commonly used notes. As shown in Table I, a note with an empty circle represents a relative duration twice as long as a note with a filled circle. The concept of relative duration also applies to rests, which denote parts where no notes are played. For example, in Figure 2, only notes with a relative duration of "1" are indicated, resulting in all notes being played at the same speed. By understanding the rhythm of the notes, the trajectory control frequency of the robot arm can be implemented through the mapping process.

B. Representation of Synergy Hand Gestures

After conducting initial investigations and engaging in a trial-and-error process, we have identified three optimal poses for piano playing. Two of these poses are suitable for playing a single note, while the remaining pose is specifically designed for playing double notes simultaneously. The reason

TABLE II
SYNERGY COORDINATES OF THE THREE POSES

	Pose 1	Pose 2	Pose 3
Synergy 1	-1	1	-0.12
Synergy 2	0.5	0.8	0.62

for designing two poses suitable for playing single notes is to enable the index finger (in the case of the right robot hand) to press notes in a relatively lower pitched range, while the ring finger can be used for a relatively higher pitched range. This approach reduces the overall movement required by the robot hand, enhancing efficiency while showcasing the capability of different poses for playing single notes.

Due to the dissimilarities in the motion capabilities between the robot hand and the human hand, the available poses for the robot hand to play the piano will differ from those used by humans. To provide further clarification, Figure 3 (b) illustrates that while the human and robot hands utilize the same fingers, their overall hand postures will differ. This disparity is particularly noticeable when comparing the utilization of the index finger (P1) and the ring finger (P2R and P2L).

The poses are a result of the two synergies, and Table II shows the coordinates of the first and second synergy for the three robot hand poses.

C. *Flow of Robot Hand Movement*

An important preliminary step before hard coding the robot's intended musical performance is to establish a mapping between the poses and their corresponding notes. This mapping ensures that the robot accurately plays the desired notes in the correct sequence. The mapping of poses is demonstrated using the song "Gong Xi Gong Xi" (a very popular traditional Chinese New Year Greeting Song) in Figure 3 (a) as an example. The breakdown of the mapping is as follows:

- The blue numbers correspond to specific finger positions in human piano playing. In this figure, the thumb is denoted by "1", the index finger by "2", the middle finger by "3", the ring finger by "4", and the little finger by "5".
- The red words correspond to the poses in robot piano playing.

As observed, a human hand has the advantage of utilizing all five fingers to their full extent, making it significantly more dexterous compared to a robot hand. However, for the robot, we would still like to utilize as many fingers as possible to allow the piano playing process to be more efficient.

Hence, by focusing on the right hand, whenever two consecutive single notes span an interval of a 3rd, which means they are 3 notes apart, we will change the pose from P1 to P2R (assuming the previous note has a lower pitch than the following note), or from P2R to P1 (in the opposite case). The change in poses at the left hand, are also designed due to this reason. The arm movement required for playing the notes can be minimized. This means that the robot hand gesture can take the dominant role in executing the movement, reducing the overall movement on the arm.

TABLE III
CONTROL ALGORITHM

```
def reset_program()
def starting_joint_angles() #Move dual arms
starting_joint_angles = [0.0, 0.0, 0.0]
return starting_joint_angles
def read_input():
poses = read_poses()
notes = read_notes()
rhythm = read_rhythm()
def system_control() # Coordination and Operation
def end_program()
while remaining_notes >0 do:
read_input()
system_control()
return
end_program()
```

This approach aligns with the preferred technique used by human pianists as well. When playing the piano, skilled human performers tend to optimize their hand and arm movements to minimize unnecessary motion on their arms, since controlling finger movement can be done more delicately compared to arm movement. By adapting this strategy to robotic piano playing, the aim is to achieve smoother and more natural-looking performances, closely resembling those of human pianists.

D. *Control Algorithm*

The program starts by resetting itself. Next, it moves the manipulator, the Franka dual arms, to the starting position. This step ensures that the manipulator is in the correct initial state for the subsequent actions. The program then reads an input consisting of three components: poses, notes, and rhythm. These inputs define specific parameters for the manipulator's movements and actions, which are already predefined as mentioned in part B. above.

After reading the input, the program moves on to system control, involving coordinating and managing the overall operation of the system. Finally, the program reaches the end of the file, indicating that it has completed all the necessary actions or instructions. At this point, the program resets itself again, potentially preparing for another round of execution or awaiting further input. The pseudocode in Table III simplifies what has been mentioned.

V. EXPERIMENTS

A. *Experimental Setup*

The CURI humanoid robot was utilized as an experimental platform. Since only three distinct hand-playing gestures (Sec. IV-B) are considered in this work, the musical piece "Gong Xi Gong Xi", which involves playing notes P1, P2, and P3, was adopted to evaluate the performance of the proposed approach. In addition, the experiment was performed using a CASIO CTK-2 electronic keyboard. All

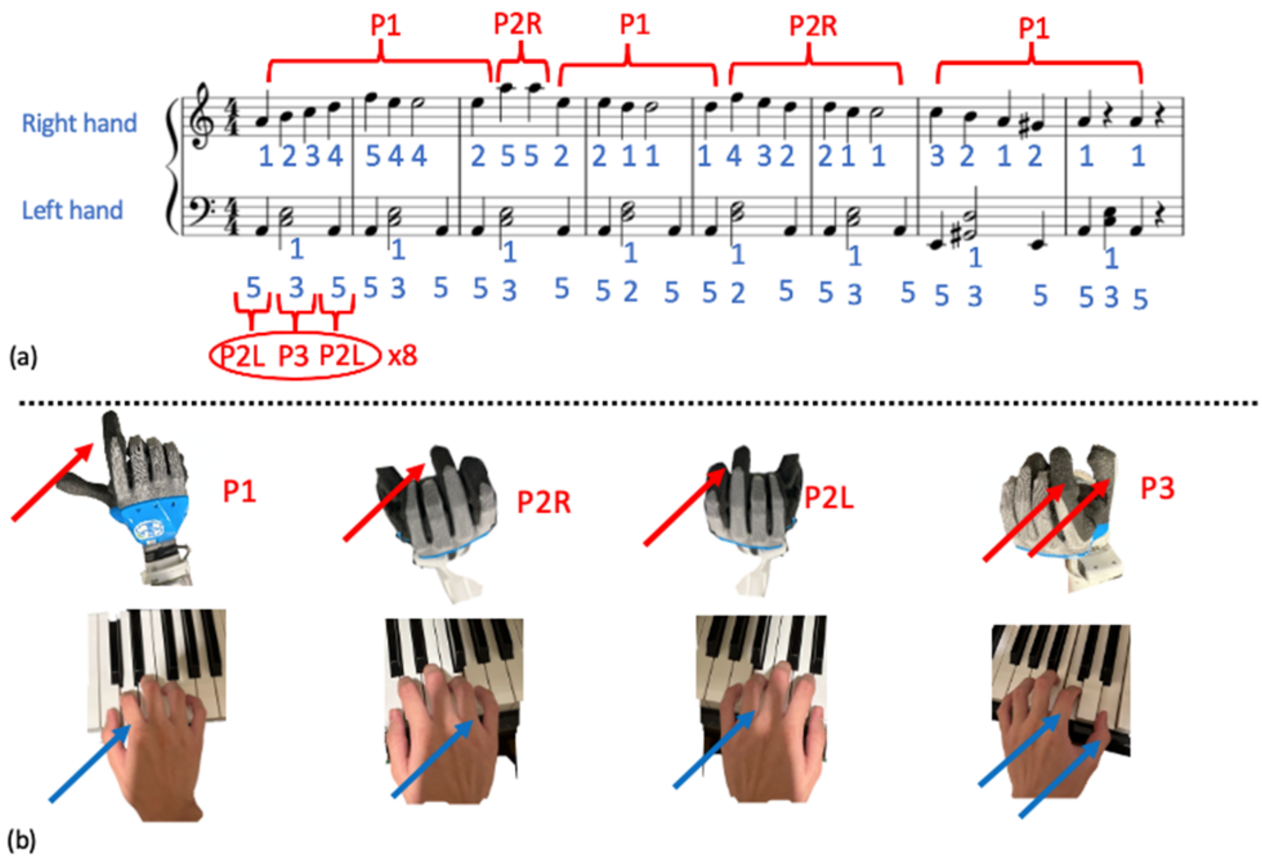

Fig. 3. (a) Comparision between human gesture and synergy-based robot hand representation for piano playing of "Gong Xi Gong Xi". (b) Comparison between the poses of hand human gestures and synergy-based robot hand representation. The blue arrow indicates the specific human finger that is pressing the keyboard. P1 presses the piano using the index finger; Right hand of pose 2 (P2R) and left hand of pose 2 (P2L) presses the piano using the ring finger; P3 presses the piano using both the thumb and index finger.

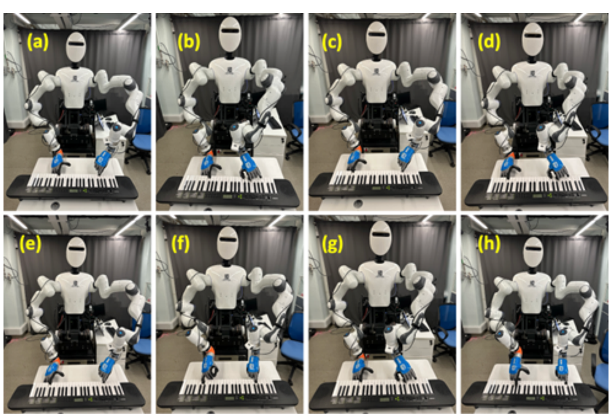

Fig. 4. Robot hand gestures and arm postures at eight selected key notes that require the coordination of both hands.

computational algorithms were executed on a graphics workstation equipped with a single Nvidia RTX 4090 GPU.

B. Evaluation Metrics

The key goal of the experiment is to evaluate the repeating playing performance of the proposed approach. Therefore, the following three quantitative metrics are utilized:

- **Playing Precision** quantifies the repeating accuracy of the proposed approach. It is calculated by the ratio of correctly played notes to the total number of notes executed by the robot.
- **Playing Recall** evaluates the repeating completeness of the proposed approach, which calculates the ratio of true positive notes to the total number of actual positive notes.
- **F1 score** combines precision and recall into a single value by calculating the harmonic mean. It serves as an effective measure for evaluating the overall playing performance of our approach.

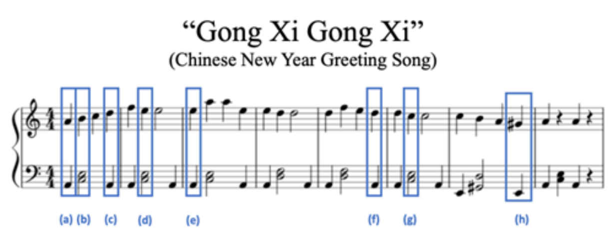

Fig. 5. The corresponding notes from Figure 4 on the music score.

C. Real Robot Experiments

The experimental procedure encompasses two main steps: 1) Human playing demonstration: Initially, an expert human pianist is required to play the entire score. The pianist's hand gestures and positions for each note are recorded during this phase. These recordings serve as the ground truth data for later comparison. 2) Robot playing repetition: In this step, the humanoid pianist replicates the previously recorded performance. The process begins with parsing the type, position, and rhythm of each note from the score, which dictates the required gestures and positions for the robot's hand, as well as the velocity of the robot's arm movements. Each note type is represented using synergy coordinates to mimic the human playing gestures closely. Following the interpretation of these parameters, the robot is programmed to execute the score using the proposed method. A note is considered successfully played if the robot precisely reaches the correct position and produces the accurate sound by striking the piano keys appropriately. Finally, the average note-playing precision, recall, and F1 score are calculated across 10 complete trials.

Table IV illustrates the experiment results, showing an average precision of 0.85, a recall of 0.66, and an F1 score of 0.74 across four notes, validating the effectiveness and accuracy of our approach. A notable observation from the results is the comparatively lower playing precision for note P2. This discrepancy is attributed to the mechanical characteristics of the Softhand2's pinky finger, which is primarily responsible for playing this note. Due to its lower stiffness, the pinky cannot exert sufficient force to consistently produce the correct sound upon key contact, adversely affecting playing precision. Additionally, qualitative results presented in Figure 4 highlight the practicability of our approach for playing eight key notes. This achievement is significant as it involves the coordinated movements of both arms and hands. It is noteworthy that such performance levels were attained using only two synergy control variables per hand, which underscores the efficiency and effectiveness of our synergy-based gesture representation in facilitating dexterous piano playing.

TABLE IV
REAL ROBOT EXPERIMENT RESULTS

Pose	Precision	Recall	F1 Score
P1	0.91	0.64	0.75
P2R	0.81	0.67	0.73
P2L	0.82	0.65	0.72
P3	0.87	0.69	0.77

VI. CONCLUSION

In this study, we explored the feasibility of employing a humanoid robot with a synergy-based robot hand for piano playing. Through mapping specific robot hand poses to notes on the piano keyboard and implementing the flow of movement through program code, we demonstrated that our system is capable of playing selected melodies with remarkable precision and coordination within 4 octaves of keys when utilizing both left and right hands. Evaluation of the system's performance using commonly used metrics like precision, recall, and F1 score showed that the synergy-based approach achieved reasonably high scores, especially for poses like P1 and P3. While the constraints of the robot hand pose fewer degrees of freedom than the human hand, we optimized movement efficiency by changing poses strategically. Overall, this work showcased the potential of synergy-based dexterous manipulation for musical applications like piano playing. However, we only consider three synergy-based playing gestures, which are insufficient for playing more complex piano scores. Future work will focus on expanding the hand gesture repertoire, improving hand synchronization, and enhancing musical expressiveness through dynamics and timing variations.

REFERENCES

[1] A. M. Batula and Y. E. Kim, "Development of a mini-humanoid pianist," in *2010 10th IEEE-RAS International Conference on Humanoid Robots*, 2010, pp. 192–197.

[2] M. Santello, M. Bianchi, M. Gabiccini, E. Ricciardi, G. Salvietti, D. Prattichizzo, M. Ernst, A. Moscatelli, H. Jörntell, A. M. Kappers *et al.*, "Hand synergies: Integration of robotics and neuroscience for understanding the control of biological and artificial hands," *Physics of life reviews*, vol. 17, pp. 1–23, 2016.

[3] R. Lin, "Ke tanzou jianpan yinyue jixieren [keyboard-playing robot]," 2010. [Online]. Available: https://www.tairoa.org.tw/school/PDF/2010/31.pdf

[4] K. Zakka, P. Wu, L. Smith, N. Gileadi, T. Howell, X. B. Peng, S. Singh, Y. Tassa, P. Florence, A. Zeng *et al.*, "Robopianist: Dexterous piano playing with deep reinforcement learning," *arXiv preprint arXiv:2304.04150*, 2023.

[5] L. Scimeca, C. Ng, and F. Iida, "Gaussian process inference modelling of dynamic robot control for expressive piano playing," *Plos one*, vol. 15, no. 8, p. e0237826, 2020.

[6] J. Hughes, P. Maiolino, and F. Iida, "An anthropomorphic soft skeleton hand exploiting conditional models for piano playing," *Science Robotics*, vol. 3, no. 25, p. eaau3098, 2018.

[7] M. Santello, G. Baud-Bovy, and H. Jörntell, "Neural bases of hand synergies," *Frontiers in computational neuroscience*, vol. 7, p. 23, 2013.

[8] "Franka research 3," n.d. [Online]. Available: https://store.clearpathrobotics.com/products/franka-research-3

[9] "qb softhand2 research," Sep 2023. [Online]. Available: https://qbrobotics.com/product/qb-softhand-2-research/

[10] G. Field, "There are multiple of each letter on a piano keyboard, so how do you know which one to play while reading sheet music?" 2021. [Online]. Available: https://www.quora.com/There-are-multiple-of-each-letter-on-a-piano-keyboard-so-how-do-you-know-which-one-to-play-while-reading-sheet-music

[11] H. Wang, T. Nonaka, A. Abdulali, and F. Iida, "Coordinating upper limbs for octave playing on the piano via neuro-musculoskeletal modeling," *Bioinspiration & Biomimetics*, vol. 18, no. 6, p. 066009, 2023.

[12] "Franka control interface (fci)," 2023. [Online]. Available: https://frankaemika.github.io/docs/

[13] "Music theory 101 - creating music literacy one step at a time." [Online]. Available: https://musictheory101.commons.gc.cuny.edu/rhythm/

[14] "General manual," 2023. [Online]. Available: https://qbrobotics.com/wp-content/uploads/2023/01/qbSoftHand2-Research-User-manual-GENERAL-4.pdf

[15] "Ros – robot operating system," 2021. [Online]. Available: https://www.ros.org/