

# Synchronize Dual Hands for Physics-Based Dexterous Guitar Playing

PEI XU, Stanford University, USA

RUOCHENG WANG, Stanford University, USA



Fig. 1. Examples demonstrating our approach to physics-based bimanual control for guitar playing with synchronized dual-hand motions. Yellow strings indicate the ones correctly pressed by the left hand and ready for the right hand to play, while green denotes those already played correctly. Policies trained using our approach can effectively coordinate the behaviors of two hands, and perform songs across diverse genres, tempos and playing techniques. From left to right, the performed songs are: *Snow (Hey Oh)* by Red Hot Chili Peppers, *Summer* by Joe Hisaishi and *Spring Festival Overture* by Huanzhi Li.

We present a novel approach to synthesize dexterous motions for physically simulated hands in tasks that require coordination between the control of two hands with high temporal precision. Instead of directly learning a joint policy to control two hands, our approach performs bimanual control through cooperative learning where each hand is treated as an individual agent. The individual policies for each hand are first trained separately, and then synchronized through latent space manipulation in a centralized environment to serve as a joint policy for two-hand control. By doing so, we avoid directly performing policy learning in the joint state-action space of two hands with higher dimensions, greatly improving the overall training efficiency. We demonstrate the effectiveness of our proposed approach in the challenging guitar-playing task. The virtual guitarist trained by our approach can synthesize motions from unstructured reference data of general guitar-playing practice motions, and accurately play diverse rhythms with complex chord pressing and string picking patterns based on the input guitar tabs that do not exist in the references. Along with this paper, we provide the motion capture data that we collected as the reference for policy training. Code is available at: <https://pei-xu.github.io/guitar/>.

**CCS Concepts:** • Computing methodologies → Animation; Physical simulation; Reinforcement learning.

**Additional Key Words and Phrases:** character animation, hand animation, physics-based control, dexterous control, motion synthesis, multi-agent reinforcement learning, cooperative learning, motion capture dataset

## ACM Reference Format:

Pei Xu and Ruocheng Wang. 2024. Synchronize Dual Hands for Physics-Based Dexterous Guitar Playing. In *SIGGRAPH Asia 2024 Conference Papers (SA Conference Papers '24), December 3–6, 2024, Tokyo, Japan*. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3680528.3687692>

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*SA Conference Papers '24, December 3–6, 2024, Tokyo, Japan*

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-1131-2/24/12...\$15.00  
<https://doi.org/10.1145/3680528.3687692>

## 1 INTRODUCTION

Facilitated by the advancements in computer graphics and VR/AR techniques, virtual music performance has emerged in recent years to provide a better immersive experience of live music [Skreinig et al. 2023]. However, due to the high requirement of control precision, it is a challenge to effectively control virtually simulated characters to interact with musical instruments in a human-like and physically plausible way. Solutions relying on motion capture recordings to reproduce motions virtually, though guaranteeing the realism of the animation, are typically costly and time-consuming, and usually unable to synthesize motions for unseen songs or rhythms. In this paper, we present an approach to synthesize motions for physically animated hands and fingers performing novel music virtually, specially, using the instrument of guitar.

The guitar is a popular instrument in daily life. However, mastering the guitar is a difficult task which needs a long time of practice, as it requires dexterity and coordination in both hands to fret chords (pressing multiple strings on the guitar neck), strum strings and execute intricate fingerpicking patterns, etc. During guitar performance, two hands undertake heterogeneous tasks interacting with strings but have to coordinate to perform temporally synchronized behaviors with exceptional spatial precision. To produce desired sounds, the left hand needs to fret (press at or release from) strings rapidly in accordance with the music rhythm, while the right hand must pluck the strings when they are in the correct pressed or released states. Similarly, some playing techniques, e.g. natural harmonics, require the left hand to wait for the right hand and release a pressed string right after it is played. Previous work explored approaches to string fretting [Heijink and Meulenbroek 2002; Radisavljevic and Driessens 2004; Wortman and Smith 2021] relying on heuristic rules to find out finger placements, but ignores the dynamics of real human hands in the temporal context during guitar playing. Our work focuses on data-driven approaches to generate human-like fretting motions that are physically plausible, exhibiting the grace and dexterity of human guitarists. Further, to provide a complete guitar-playing experience, we introduce bimanual control to generate motions where

the fretting and strumming hands play a given song on a virtual guitar at the same time with sounds generated through post-processing based on the two hands' dynamics.

In computer graphics, previous works of hand motion synthesis mostly focus on object grasping and/or manipulating [Andrychowicz et al. 2020; Xie et al. 2023; Yang et al. 2022; Zhang et al. 2021; Zhao et al. 2013], or hand gesture generation [Kucherenko et al. 2020; Liang et al. 2022; Qian et al. 2021; Zhang et al. 2024]. Effective bimanual control with synchronized but heterogeneous tasks, like guitar playing, is under-investigated. Typically, bimanual control can be performed by considering two hands as a unified agent. However, training one joint policy for two hands would exponentially increase the dimension of the state-action space, compared with the case of having only one hand. Meanwhile, controlling two hands in a centralized environment would introduce more computation costs for simulation.

Instead of regarding two hands as a unified agent and performing policy training jointly, we consider guitar playing as a cooperative task where the left and right hands are treated as individual agents. To improve the learning efficiency, our approach decouples the two-hand control first by training the policies for individual hands separately in decentralized environments, and then quickly adapts single-hand policies to generate synchronized behaviors in a centralized environment by manipulating the latent spaces of the two individual policies. Our approach can improve the overall learning efficiency by reducing the training required for cooperative learning in centralized environments.

While string fretting and playing are goal-driven tasks during guitar playing, we perform physics-based control with additional imitation learning to ensure the naturalness and physical plausibility of the generated motions in such contact-rich tasks. We captured a set of hand motions approximately one hour long from a professional guitarist. All the motions are general guitar-playing techniques rather than the tracking of playing any specific songs. While taking two short clips of scaling and strumming motions as the references for imitation learning, our approach shows excellent performance in synthesizing those unstructured reference motions to generate natural guitar-playing motions with automatic fingering inference for novel string fretting and playing patterns.

We evaluate our approach both qualitatively and quantitatively. As shown in the experiment section, while learning high-quality single-hand control, our approach can efficiently synchronize single-hand policies and effectively coordinate two hands' behaviors to provide dexterous motions with high precision both temporally and spatially. Additional ablation studies in the supplementary materials support the proposed training scheme for bimanual control along with highlighting the potential of our approach to apply to tasks other than guitar playing.

## 2 RELATED WORKS

Traditional methods for physics-based control typically rely on trajectory optimization and/or human-designed heuristic rules for control [Chen et al. 2023b; Liu 2008, 2009; Mordatch et al. 2012; Wang et al. 2013; Ye and Liu 2012]. Early exploration of imitation learning using pre-collected mocap data [Kry and Pai 2006; Pollard

and Zordan 2005; Zhao et al. 2013] achieved impressive results to generate human-like motions. In recent years, with the advancement of deep learning techniques, reinforcement learning based imitation learning [Ling et al. 2020; Merel et al. 2017; Peng et al. 2018, 2022, 2021; Won et al. 2020; Xu and Karamouzas 2021; Yao et al. 2022, 2023] has drawn lots of attention and become a popular approach for physics-based character control. Our approach follows the literature and focuses on physics-based dexterous bimanual control using reinforcement learning.

### 2.1 Dexterous Control for Physically Simulated Hands

Dexterous control has attracted significant attention due to its complexity and frequent occurrence in everyday situations. Typical dexterous control tasks can generally be divided into two main categories: object grasping and/or relocation [Liu 2009; Xie et al. 2023; Zhao et al. 2013] and in-hand manipulation [Andrychowicz et al. 2020; Yang et al. 2022; Zhang et al. 2021]. The common scenarios of these tasks do not have a high requirement of temporal precision, and often only need one hand. In contrast, our guitar-playing task requires precise control of two hands simultaneously.

Generating human-like instrument-playing motions based on given music has been studied extensively in recent years [Chen et al. 2023a; Hirata et al. 2022; Kao and Su 2020; Li et al. 2018; Liu et al. 2020; Shlizerman et al. 2018]. Early work [ElKoura and Singh 2003] considers motion generation for guitar playing. However, they only focus on the left hand's fretting motions rather than the coordinated motions of two hands. Besides, their method is kinematics-based, ignoring the constraints of finger transitioning under the temporal context and the complicated collision and contact states between fingers and those between the fretboard and fingers.

Our goal in this paper is to generate physically plausible two-hand motions for guitar playing. Instead of using heuristic rules to perform control and motion generation for virtual musical instrument playing [ElKoura and Singh 2003; Zakka et al. 2023; Zhu et al. 2013], we employ imitation learning plus reinforcement learning to perform motion synthesis from unstructured reference data [Xu et al. 2023a], accounting for the dexterity and grace of human guitarists, and rely on the control policy to perform fingering inference through exploration on its own without needing to explicitly design any heuristic rules for motion generation.

### 2.2 Hand Motion Datasets

Currently, most hand motion datasets only provide motions involving object grasping or manipulation [Chao et al. 2021; Fan et al. 2023; Taheri et al. 2020]. Only a few datasets contain motions of playing musical instruments like violin and piano [Grauman et al. 2024; Simon et al. 2017]. To our knowledge, there is still a lack of publicly available datasets containing high-quality motions of guitar playing. Along with this paper, we provide the guitar-playing motions that we captured from a professional guitarist. The whole dataset is around 1 hour long, including the practice of various guitar-playing techniques. From the dataset, we take two clips of scaling and strumming motions for our policy training. We refer to the supplementary materials for the details of our dataset and the motion capture setup.

### 3 OVERVIEW

An overview of our proposed policy synchronization system for learning physics-based bimanual control is shown in Fig. 2. To avoid directly performing training in a centralized environment with two hands, in this system, we treat the left and right hands as independent agents, with their control policies trained separately. The two single-hand control policies are finally put into a centralized environment for cooperative learning to generate synchronized two-hand motions. Instead of directly fine-tuning the two pre-trained single-hand policies, we introduce a synchronizer to quickly coordinate two policies' behaviors via manipulating the latent spaces  $\mathbf{z}_t^L$  and  $\mathbf{z}_t^R$  for the left and right hand respectively.

For single-hand policy training, our approach takes the multi-objective reinforcement learning framework [Xu et al. 2023a], considering both the motion imitation and task objectives of fret pressing for the left hand and string picking for the right hand. The single-hand policies are trained for general purposes using music notes extracted from real songs and those generated procedurally for the left and right hands respectively. For policy synchronization, we extend the policy adaptation strategy from [Xu et al. 2023b] and introduce a synchronizer to coordinate two control policies simultaneously through latent space manipulation to generate guitar-playing motions for specific target songs. To provide a better understanding of the whole system, we will first elaborate on the synchronization process in Section 4, and then detail the implementation of policy training for the dexterous dual-hand guitar playing task in Section 5.

### 4 POLICY SYNCHRONIZATION

Under the setup of reinforcement learning, A two-agent cooperative task can generally be considered an optimization problem to maximize the cumulative rewards associated with two agents jointly, i.e.,  $\max \mathbb{E}_{\tau \sim p(\tau)} [\sum_t \gamma^{t-1} \mathcal{R}_t]$  where  $\tau = \{\mathbf{o}_t^L, \mathbf{o}_t^R, \mathbf{a}_t^L, \mathbf{a}_t^R\}_t$  is the state-action trajectory of the two agents denoted by  $L$  and  $R$  respectively,  $\gamma$  is a discount factor,  $\mathcal{R}_t$  is a scalar reward to evaluate the two agents' joint performance at the time step  $t$ . A joint action policy  $\pi(\mathbf{a}_t^L, \mathbf{a}_t^R | \mathbf{o}_t^L, \mathbf{o}_t^R)$  can be achieved in a centralized environment by the same way that we train single-agent policies.

Inspired by how humans perform cooperative tasks, we propose a two-step training scheme to perform two-agent cooperative learning efficiently. Our scheme first decouples the cooperative task into a set of sub-tasks for individual agents and performs policy training independently for each agent, which is like that humans learn basic work skills before being involved in a cooperative task; and then the two individual policies,  $\pi_L(\mathbf{a}_t^L | \mathbf{o}_t^L)$  and  $\pi_R(\mathbf{a}_t^R | \mathbf{o}_t^R)$ , are adapted in a centralized environment to serve as a joint policy  $\pi(\mathbf{a}_t^L, \mathbf{a}_t^R | \mathbf{o}_t^L, \mathbf{o}_t^R)$  providing synchronized cooperative behaviors for two-agent control. Through this setting, we avoid directly performing optimization in the joint state-action space and, meanwhile, improve the overall training efficiency by preventing directly performing cooperative learning for two agents lacking basic skills for cooperation.

As we assume that the individual agents are already trained well during the first individual training phase and have the potential to provide needed behaviors for cooperation, our goal during policy synchronization is to quickly coordinate individual policies' behavior patterns conditionally for cooperation rather than letting them

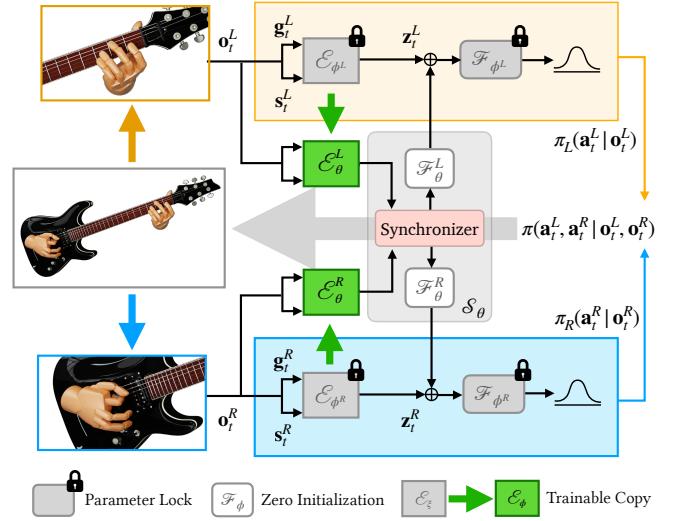


Fig. 2. Overview of the proposed system synchronizing dual policies for dexterous guitar playing with two hands. Our system performs two-hand policy training in two steps. First, we decentralize the control of two hands, and train the left-hand policy for fret pressing (orange box) and the right-hand policy for string picking (blue box) independently in a decentralized manner. Then, we lock the previously trained single-hand policies, and introduce a synchronizer to coordinate the behaviors of single-hand policies in a centralized training environment to obtain a joint policy for two-hand control. The synchronization is achieved quickly by modifying single-hand policies' behavior patterns through latent space manipulation.

learn new behaviors from scratch. Therefore, we do not directly fine-tune the networks of the pre-trained individual policies. Instead, as shown in Fig. 2 we introduce a synchronizer to take the input from both individual policies and conditionally coordinate their behaviors by outputting offsets to the two policies' latents. During synchronization, all parameters in the pre-trained policy networks are locked, and we only optimize the synchronizer-related parameters  $\theta$ . Though we show the same network structure for individual policies,  $\pi_L(\mathbf{a}_t^L | \mathbf{o}_t^L)$  and  $\pi_R(\mathbf{a}_t^R | \mathbf{o}_t^R)$ , in Fig. 2, there is no requirement for them to be identical in practice.

Following the previous literature [Xu et al. 2023b], the synchronizer in our implementation is built using a multilayer perceptron (MLP) architecture but with the last fully-connected layers initialized by zero ( $\mathcal{F}_\phi^L$  and  $\mathcal{F}_\phi^R$  in Fig. 2). This means that the synchronizer will not change the behavior of individual policies when synchronization training starts. The policy exploration in the synchronization phase, therefore, will start not randomly but from the trajectory generated by the pre-trained individual policies, and thus would be more efficient. Additionally, we take advantage of the pre-trained individual policies, and initialize the synchronizer's state encoders,  $\mathcal{E}_\theta^L$  and  $\mathcal{E}_\theta^R$ , by duplicating parameters from the corresponding state encoders in the pre-trained policies ( $\mathcal{E}_{\phi^L}$  and  $\mathcal{E}_{\phi^R}$  respectively in Fig. 2). The process of latent space manipulation can be written as:

$$\mathbf{z}_t^L = \mathcal{E}_{\phi^L}(\mathbf{o}_t^L) + \Delta \mathbf{z}_t^L, \text{ and } \mathbf{z}_t^R = \mathcal{E}_{\phi^R}(\mathbf{o}_t^R) + \Delta \mathbf{z}_t^R \quad (1)$$

where  $\mathbf{z}_t^L$  and  $\mathbf{z}_t^R$  are the latents of the two policies  $\pi_L$  and  $\pi_R$ , and

$$\Delta \mathbf{z}_t^L, \Delta \mathbf{z}_t^R = \mathcal{S}_\theta \left( \text{CONCAT} \left( \mathcal{E}_\theta^L(\mathbf{o}_t^L), \mathcal{E}_\theta^R(\mathbf{o}_t^R) \right) \right). \quad (2)$$

During the synchronization phase, the joint policy can be updated by optimizing the synchronizer with parameter  $\theta$  through general reinforcement learning algorithms.

## 5 LET'S PLAY GUITAR

In Fig. 3, we show the profile of a common guitar model. During guitar playing, typically, the left hand moves along the fretboard to press one or more target strings at the positions of different string-fret combinations, while the right hand interacts with the strings using fingers and/or through a plectrum called *pick*. In the following, we use the terminology *chords* to describe cases where more than one string needs to be pressed by the left hand at the same time; and for the right hand, we focus only on the case where a pick held between the thumb and index is used to interact with the strings.

### 5.1 Left-Hand Policy for Fret Pressing

Guitar tablature (tab), as shown in Fig. 3, is a common visual representation of music notes for guitar playing, which indicates the target strings for picking and the target fret at which a string should be pressed before being picked while 0 means an open string that should be picked without any fret being pressed. We take the format of tab as the input goal state to the left-hand policy, but using -1 to represent open strings and 0 to represent mute strings. We consider a horizon of five future notes, and along with each note, a timer variable is employed to indicate the remaining duration of that note in terms of the number of simulation frames. This results in a goal state vector  $\mathbf{g}_t^L \in \mathbb{R}^{5 \times 7}$ . The pose state  $\mathbf{s}_t$  is composed of the position, orientation, and linear and angular velocities of each link in the hand model related to the guitar. We take a 2-frame historical observation, which gives us a state vector  $\mathbf{s}_t \in \mathbb{R}^{2 \times 16 \times 13}$ .

Fingering strategies for fret pressing involve the chord patterns of the current and future notes and depend on the biological structure (movement constraints, energy consumption, etc.) of human hands. Besides using fingertips to press single strings, some chords need one finger to press multiple strings or need multiple fingers to press different strings at the same fret simultaneously. Those challenges make the fingering problem intricate. In our implementation, we do not explicitly provide additional fingering inference, but rely on the control policy itself to explore fret-pressing strategies.

Since every string has its own target pressing condition at each time step  $t$ , we consider the fret-pressing task as a multi-objective problem, where the target pressing condition of each string is regarded as an objective. Given the three possible conditions, we define the goal-driven reward function for each string  $i$  as

$$r_i^k = \begin{cases} r_i^k & \text{if string } i \text{ should be pressed at fret } k \text{ for picking,} \\ r_i^0 & \text{if string } i \text{ should not be pressed (open string),} \\ r_i^{-1} & \text{otherwise (mute string).} \end{cases} \quad (3)$$

To encourage fret pressing,  $r_i^k$  is defined based on the *minimal* distance from the target fret-string position  $\mathbf{p}_{i,k}$  and all *available*

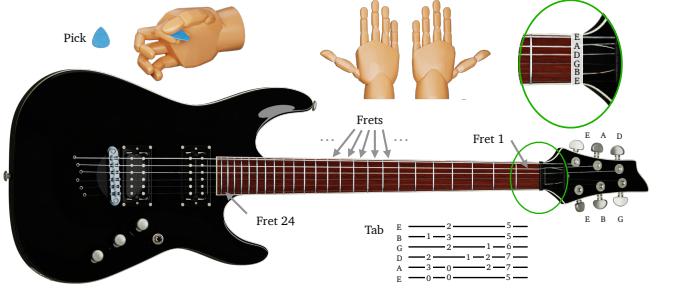


Fig. 3. Profile of the simulated guitar and hand models. Our guitar model is left-handed. It has six strings with twenty-four frets and is played with a pick. Given a common guitar chord spanning at most four frets plus two additional conditions – open string (being picked without fret pressing) and mute string (not being picked to generate any sound) – for each string, there are nearly 900,000 chord combinations theoretically, which makes it difficult to be fully mastered. Each of our hand models has 16 links with 27 degrees of freedom (DoFs), where the wrist joint has 6 DoFs, the MCP joints have 2 DoFs with the exception that the thumb MCP having 3, and all the other finger joints have 1 DoF. The simulated guitar has a scale length of 24.5 inches (around 0.62m) in our implementation. In this work, we only consider the control of hands and assume that the guitar is fixed in the 3D space. We can obtain right-handed guitar-playing motions by mirroring the setup of the hand and guitar.

finger parts at the time step  $t$ :

$$r_i^k = 0.8 \exp(-1000d_{i,k}^2) + 0.2 \exp(-30d_{i,k}^2) \quad (4)$$

where  $d_{i,k} = \min_j \|\mathbf{p}_{j \rightarrow i,k} - \mathbf{p}_{i,k}\|$ ,  $\mathbf{p}_{i,k}$  is the intersection position of string  $i$  and the middle line of fret  $k$ , and  $\mathbf{p}_{j \rightarrow i,k}$  is the closest point on an *available* finger part  $j$  to  $\mathbf{p}_{i,k}$ . Each finger is considered to have three cylinder-shaped parts between its MCP, PIP, DIP, and tip. The hand model has 12 finger parts totally, which can be used for string fretting, excluding the thumb. The closest point  $\mathbf{p}_{j \rightarrow i,k}$  can be found by calculating the distance between the point  $\mathbf{p}_{i,k}$  to the axis segment of the finger-part cylinder  $j$ .

In practice, when using a finger part other than tips to press a string, the finger typically has to lie down on the fret to perform a strong pressing. Hence, we consider a finger part between MCP and PIP or that between PIP and DIP *available* for string pressing only if the angle between it and the fretboard surface is less than  $5^\circ$ . In challenging cases where strings at multiple frets need to be pressed simultaneously, one finger typically would not be used to perform pressing across multiple frets or cross over other fingers in order to generate a clear sound. Therefore, we additionally decide if a finger's three parts are *available* for pressing at strings on a target fret based on the discrepancy between the finger order and the order of the target fret, as demonstrated in Fig. 4. We compute  $r_i^k$  only considering *available* finger parts.

Since the fretboard is around 0.45m long with fret widths varying from 0.01m to 0.035m,  $r_i^k$  in Eq. 4 is designed as the sum of two exponential functions, where the former one with a higher weight ensures that the function is sensitive enough when  $d_{i,k}$  is around 0.01m, while the later one prevents the function falling into the saturation range when  $d_{i,k}$  is near 0.45m.

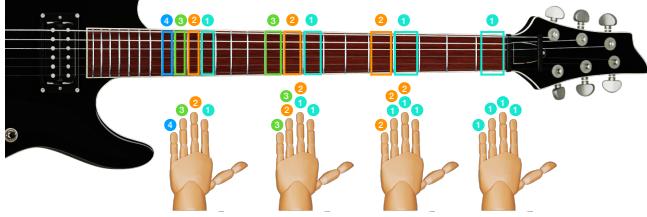


Fig. 4. Demonstrations of finger availability for fret pressing. We allow multiple fingers to press at one fret but do not allow fingers to cross over. Each finger can press at no more than one fret, and at most four frets could be pressed simultaneously. Here we show the four possible cases with the one (right) to four (left) target frets that need to be pressed. Target frets and the corresponding available fingers that can be used to press at the target frets are labeled using the same number.

For an open string that should be picked without being pressed,  $r_i^0$  in Eq. 3 induces fingers not to touch the string. It is defined based on the minimal distance between the string and all fingers:

$$r_i^0 = \min \{ (d_i/0.007)^2, 1 \} \quad (5)$$

where the  $d_i$  is the shortest distance between string  $i$  and all 12 finger parts. The constant scalar 0.007 is chosen since our hand model has fingers with a radius of around 0.006m. Therefore,  $d_i/0.007$  indeed is a normalized distance, and guarantees that string  $i$  is not touched by any finger, when it is equal to or greater than 1.

The third case in Eq. 3 is for a mute string, which is assumed not to generate any sound during the period of a note. As mute strings will not be picked, their pressing conditions are not strictly constrained. However, to produce clear sounds, mute strings theoretically should be touched slightly (not pressed) to mute their vibration if they were picked and still vibrating. Here, we use a loose constraint

$$r_i^{-1} = 0.9 + 0.1r_i^0 \quad (6)$$

to encourage but not force fingers to keep away from the string.

To take into account the left hand's overall performance, two common terms are added to  $r_i$  for all the six string-based goal-driven objectives. The final reward for each objective  $i$  is

$$r_{t,i}^L = 0.8r_i + 0.2r_{\text{correct}}^L - 0.05r_{\text{energy}}^L \quad (7)$$

where  $r_{\text{correct}}^L = 1$  if all strings are pressed at the target frets correctly or 0 otherwise, and  $r_{\text{energy}}^L$  is a term measuring the energy consumption based on the linear velocities of wrist and fingers:

$$r_{\text{energy}}^L = \exp \left( - \left( \|v_w\| + 0.1 \sum_f \|v_f\| \right)^2 \right) \quad (8)$$

where  $v_f$  is the linear velocity of each fingertip locally related to the wrist, and  $v_w$  is the velocity of the wrist in the global space.

To generate natural motions, we also employ two imitation objectives. Following the previous literature of composite motion learning [Xu et al. 2023a], we decouple the left-hand motion into (1) the wrist-thumb motion related to the guitar and (2) the local motions of the other four fingers related to the wrist, and use a GAN-like architecture [Xu and Karamouzas 2021] to perform imitation learning. Such a decoupling strategy ensures that the hand can naturally hold the guitar through a correct wrist and thumb pose provided

by the reference motions, while the pose of the other four fingers is free from the guitar's pose.

Written in a general on-policy learning form, the final optimization objective for policy training is:

$$\max \mathbb{E}_t \left[ \sum_{\kappa} w_{\kappa} \bar{A}_{t,\kappa}^L \log \pi_L(a_t^L | o_t^L) \middle| \phi^L \right] \quad (9)$$

where  $\bar{A}_{t,\kappa}^L$  is the standardized advantage that is estimated according to the goal-driven or imitation objective reward  $r_{t,\kappa}^L$ , and  $w_{\kappa}$  is the weight associated with each object  $\kappa$ . We have  $w_{\kappa} = 0.15$  for each of the six goal-driven objectives of fret pressing and  $w_{\kappa} = 0.05$  for each of the two imitation objectives. Following the previous literature, we use DPPO [Schulman et al. 2017] as the backbone reinforcement learning algorithm which takes an actor-critic architecture for policy optimization.

## 5.2 Right-Hand Policy for String Picking

The right-hand policy performs string picking and thus only needs to know if a string should be picked or not. Therefore, compared to the left-hand policy, the goal state  $g_t^R$  is defined using only a binary vector for tab representation without the fretting information. During the duration of one note, a target string is expected to be picked only once. To let the policy ignore strings that are already picked, if a target string is picked, its corresponding representation in  $g_t^R$  will be set to 0 in the remaining frames of that note. While the left hand's performance of fretting at each time step  $t$  is decided only by the fretting state at that time step, the right hand's performance of picking is evaluated by the picking behavior during the duration of each note. To provide better state-value estimation, the critic network additionally takes a binary vector indicating if each string is tackled wrongly throughout the current note. This leads to an extended goal state  $g_t^{R+} \in \mathbb{R}^{5 \times 7+6}$  for the critic network. Such an implementation is crucial for efficient policy training.

To generate sounds in unison, when multiple strings need to be picked, they must be picked in rapid succession with a consistent order from either up to down or down to up. Therefore, differing from the left-hand policy that has multiple string-based goal-driven objectives, we consider the string-picking task as an integrated objective. The reward function is designed to evaluate the picking performance at the time step  $t$  or the overall performance during the duration of the current note if no target string is left:

$$r_{\text{pick}} = \begin{cases} r^+ & \text{if any string is picked at the time step } t, \\ r^- & \text{if no target string is left for picking,} \\ r^\times & \text{otherwise.} \end{cases} \quad (10)$$

$r^+$  is to evaluate the picking behavior if it happens. Since we require strings to be picked in a consistent order, a string is considered tackled wrongly (1) if the string is a non-target string but picked, (2) if the string is not picked but a string below it is picked when the picking direction is top to down, or (3) if the string is not picked but a string above it is picked when the picking direction is down to up. We define  $r^+$  based on the distance from the pick tip to the wrongly tackled string(s) and encourage the pick to move away to

the farthest wrongly tackled strings, i.e.

$$r^+ = \begin{cases} \min_i r_{d_i} & \text{if any string } i \text{ is tackled wrongly at step } t, \\ r_{\text{succ}} & \text{otherwise,} \end{cases} \quad (11)$$

where  $r_{d_i}$  is a reward defined by the shortest distance from the pick tip to the wrongly tackled string  $i$ . Here we employ a form similar to Eq. 4 to define  $r_{d_i}$  as

$$r_{d_i} = 0.175 \exp(-10000d_i^2) + 0.025 \exp(-2000d_i^2) \quad (12)$$

taking into account that the distance between the pick tip to a string during guitar playing would vary from millimeters to decimeters.  $r^+ = r_{\text{succ}} = 1$  if not any string is tackled wrongly, namely, the top-most string(s) are picked from up to down or the bottom-most string(s) are picked down to up.

$r^-$  in Eq. 10 is used to evaluate the overall picking performance during the current note. It takes effect only when all target strings have been picked or if there is not any target string in the current note for picking.  $r^-$  is defined as

$$r^- = 0.4 \min \{\min_i d_i / 0.003, 1\} + 0.1 \sum_i r_i + 0.5 \prod_i r_i \quad (13)$$

where  $i = 1, \dots, 6$  indicates a string,  $d_i$  is the shortest distance from the pick tip to the string  $i$ ,  $r_i = 1$  for a target string if it has been picked correctly or for a non-target string if it was not picked accidentally, or 0 otherwise. The term  $\prod_i r_i$  is an additional reward effecting only when all six strings are tackled correctly. Considering the string radius, here we expect the pick tip to be at least 0.003m away from any strings.

We take  $r^\times$  to entice correct picking behaviors if one or more strings need to be picked but no picking happens at the time step  $t$ :

$$r^\times = 0.2 + 2 \min\{r_{d_{\text{top}}}, r_{d_{\text{bottom}}}\} \quad (14)$$

where the subscript *top* and *bottom* indicate the top and bottom-most target strings respectively.  $r^\times$  encourages the policy to move the pick to the nearest top or bottom-most target string. At each time step  $t$ , we always have  $1 = r_{\text{succ}} > r^\times \geq 0.2 \geq \min_i r_{d_i}$ , which brings a higher priority for the policy to move to and pick strings from the top or bottom-most target one.

Besides the picking performance, we consider the contact condition of the thumb and index fingers for pick holding and the energy consumption caused by the hand and pick movement. The overall goal-driven reward for the right-hand policy training is:

$$r_t^R = r_{\text{pick}} + 0.05r_{\text{contact}} + 0.05(r_{\text{energy}}^R + r_{\text{energy}}^{\text{pick}}) \quad (15)$$

where  $r_{\text{contact}} = 1$  if the thumb and index contact with each other correctly to hold the pick or 0 otherwise;  $r_{\text{energy}}^R$  is the reward for the energy consumption of the hand's movement and defined in the same way with  $r_{\text{energy}}^L$  in Eq. 8; and  $r_{\text{energy}}^{\text{pick}}$  is a term used to produce stable movement of the pick, and is defined based on the pick's linear velocity  $\mathbf{v}_{\text{pick}}$  and acceleration  $\mathbf{a}_{\text{pick}}$ :

$$r_{\text{energy}}^{\text{pick}} = \exp(-20||\mathbf{v}_{\text{pick}}||^2) - 14||0.05\mathbf{a}_{\text{pick}}||^4. \quad (16)$$

Similar to the left-hand policy training, we rely on imitation learning to generate natural human-like motions. Nevertheless, the right-hand motion using a pick to strum strings, though not as diverse as the left-hand motion, has more subtle dynamics. As the hand sweeps up and down, fingers would show compliant motions

with the movement of the wrist. Therefore, we consider the whole right hand as a group for motion imitation. The right-hand policy is optimized using the formula of Eq. 9 but with only two objectives (one goal-driven objective  $r_t^R$  and one imitation objective) having an equal weight of 0.5.

### 5.3 Dual Hand Policy Synchronization

After achieving two single-hand policies, we perform policy synchronization using the approach depicted in Section 4. We adopt the same multi-objective learning method (cf. Eq. 9) to optimize the joint policy for specific target songs during synchronization, but update only the synchronizer-related parameter  $\theta$  rather than the pre-trained policy networks. We keep the six goal-driven objectives and two imitation objectives for the left hand. For the right hand, however, we take only the goal-driven objective of string picking without any imitation learning, as we found that the right-hand policy can effectively master picking motions without needing further training given the limited number of string-picking patterns, while the left hand still needs the reference motion to help learn natural motion poses when facing novel chord patterns. To introduce the additional cooperative condition between two hands, when computing the reward  $r_{\text{pick}}$  for the right hand, it will be regarded as there is no target (using the  $r^\times$  case in Eq. 10) unless all strings are at the expected pressing states.

## 6 EXPERIMENTS

We run experiments using IsaacGym [Makoviychuk et al. 2021] as the physical engine and taking a hand model modified from the Modular Prosthetic Limb model [Kumar and Todorov 2015]. Our hand model, as shown in Fig. 3, is skeleton-driven where each joint is modeled as a motor controlled through a PD controller, which results in an action space  $\mathbf{a}_t^L, \mathbf{a}_t^R \in \mathbb{R}^{27}$ . All simulation runs at 240Hz with the control policies running at 60Hz. Due to the complication of simulating string dynamics, all strings are modeled virtually rather than physically simulated. A string is considered pressed by a finger part if the distance between them is less than 0.006m (approximately the radius of fingers). A string is assumed to be picked if the pick tip goes through that string's virtual position from one side to the other side with a trajectory below the string.

We perform general-purpose training for the left-hand policy using notes extracted from 500 music tracks with a wide range of rhythm styles and tempos. We further augment the extracted notes online during training by randomly shifting them along the fretboard, while keeping the chord pattern unchanged, plus a stochastic tempo adjustment of  $\pm 20$  beats per minute (BPM). For the right hand, we use procedurally generated strumming patterns as the target goals for training. During synchronization, behaviors of the two pre-trained individual policies are fine-tuned using specific target songs. We take (1) a clip of scaling practice motions around 40 seconds long including two hands and (2) a clip of the right hand's strumming motion which is about 3 seconds long as the references to synthesize motions for novel music note playing instead of performing motion tracking. We refer to the supplementary materials for additional implementation and data collection details.

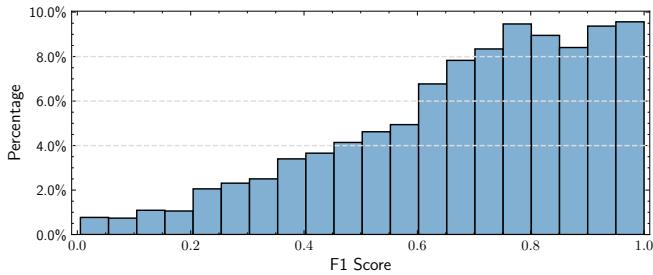


Fig. 5. Distribution of F1 scores achieved by the left-hand policy when playing chords. The test set contains 50 music tracks with 1721 unrepeated measures having 4859 chords.

## 6.1 Numeric Results

*Metrics:* We use F1 scores to evaluate the trained policies’ performance on the level of music notes. A true positive sample in our case means a string that is correctly pressed or picked during the duration of a note. Precision indicates the ratio of true positive samples to the total number of strings that are pressed or picked. Recall indicates the percentage of the correctly handled strings among the target strings of one note. For the left hand, a string is considered pressed correctly if it is pressed at the target fret continuously for at least 2/3 of the duration of the note. For the right hand, a string is considered picked correctly if it is a target string and is picked in a consistent order (cf. Section 5.2) once and only once during the note’s duration. For the joint policy of two hands, a string is treated as being handled correctly only if (1) the string is picked correctly, (2) the string is being pressed correctly when being picked, and (3) the pressing state should be kept for at least 2/3 duration of the note after the string is picked.

*Single-hand policy evaluation:* We evaluate the performance of the separately trained single-hand policies using a test set containing 50 music tracks. We remove all the repeated measures and keep 4859 notes left. The overall F1 score achieved by the left-hand policy is around 0.8. For regular notes where only one string needs to be pressed, the policy obtains almost 100% accuracy even for fast songs with a tempo as high as 150BPM, which is challenging for human players. For chord notes playing, the left-hand policy can obtain an F1 score above 0.8 for around 40% of the tested chords. The distribution of the F1 scores for chords playing is shown in Fig. 5. From the figure, we can see that there are around 20% chords for which the policy achieves an F1 score of less than 0.5. We find that the poorly performed chords are those not in or only appearing few times in the training set for the left hand. Though we employ a large set of 500 music tracks for the left-hand policy training, it is not enough to fully cover all the possible chords that may appear during guitar playing. In contrast, the right hand policy only could face a limited number of picking patterns, and achieves consistent performance with an average F1 score above 0.9 for all the tested notes. Given the sophistication of rhythm and chord generation, it would be an interesting direction for future work to design programs that can provide diverse chords with a better balanced distribution for the left-hand policy to learn.

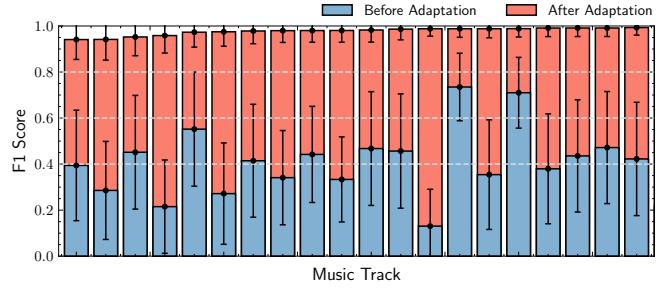


Fig. 6. F1 scores of bimanual performance on a test set of 25 music tracks before and after synchronization. We show both the mean values and the standard deviations of F1 scores evaluated on the level of music notes.

*Synchronization performance evaluation:* To evaluate the effectiveness of the proposed policy synchronization scheme, we perform policy synchronization for 25 music tracks using the pre-trained single-hand policies, and report the F1 scores before and after synchronization in Fig. 6. Due to the cooperation requirement, the F1 scores involving two hands typically are lower compared to the case when evaluating the single-hand policies independently. As shown in the figure, policy synchronization brings about a significant improvement in F1 scores, some of which are as high as 657% with an average improvement of 193%. An interesting observation is that we only perform latent space manipulation to change the policies’ behavior patterns rather than fine-tuning the whole policy networks, and the right-hand policy even does not perform any imitation learning during synchronization. However, the single-hand policies’ performance still could be improved effectively. This means that the single-hand policies were trained well during the pre-training phase to provide poses potentially needed during guitar playing. Overall, the two-hand policy can achieve almost 100% accuracy for most tested songs after synchronization. Such improvement partially is because of the fine-tuning effect during policy synchronization, but also comes from the synchronized two-hand behaviors. To further demonstrate that, we studied the change in F1 scores when some nearly perfect single-hand policies were taken for synchronization. We refer to the supplementary materials for the related analysis.

## 6.2 Qualitative Evaluation

We show the snapshots of our control policies’ live performance in Fig. 1. Policies trained by our approach can be synchronized for music with distinct tempos, genres and playing techniques, and generate visually convincing animations for novel rhythms not included in the reference motion data.

For the left hand, we highlight the control policy’s accurate fretting ability in Fig. 7. As can be seen, the hand can perform long-distance movements swiftly to fret target strings. The wrist part’s traveling distance in the 3rd case, for example, crosses 11 frets and reaches around 0.3m. Note that the maximal fret is 0.035m wide while the minimal is just around 0.01m. Such small widths request a demand for precise control spatially. In Fig. 8, we show the dexterous control of the left hand performing diverse novel chords. All chords shown in the figure do not exist in the reference motions



Fig. 7. Examples of the left-hand policy swiftly and accurately pressing strings at the target frets in the distance. Strings highlighted in green are target strings that are pressed correctly. Policies trained by our approach can perform accurate long-distance pressing using multiple fingers at the same time.



Fig. 8. Examples of the diverse chords performed by the left-hand policy. Green spots indicate the target string-fret positions while some of them are occluded by the hand due to accurate fretting.

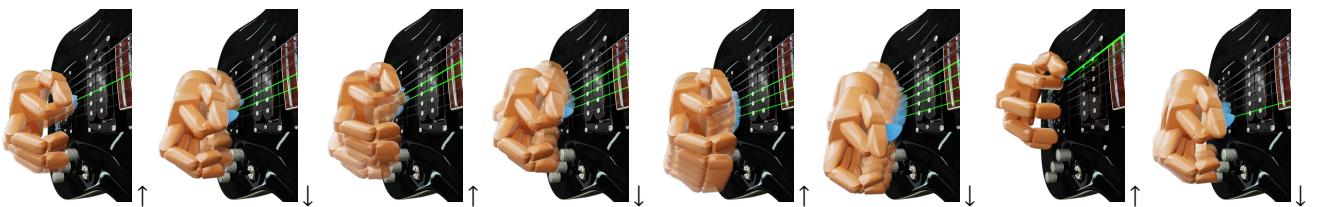


Fig. 9. Complex picking patterns during the right-hand policy playing the song *Hotel California*. While accurately picking the target strings, our policy follows an efficient trajectory to move up and down and up again consistently. The arrows indicate the picking direction.



Fig. 10. Fast string strumming to play chords. Correctly strummed target strings are shown in green. Our policy can perform strumming smoothly in rapid succession with high accuracy. The right figure shows a pair of consecutive strumming, where strings are picked from down to up and then up to down.



Fig. 11. Motions generated by policies trained using only goal-directed task rewards without motion imitation.

Fig. 12. Two-hand performance before (top) and after (bottom) policy synchronization. Red strings indicate the target strings that are picked without being correctly fretted. Yellow strings in the right-bottom figure indicate those that are correctly fretted and have not been picked.

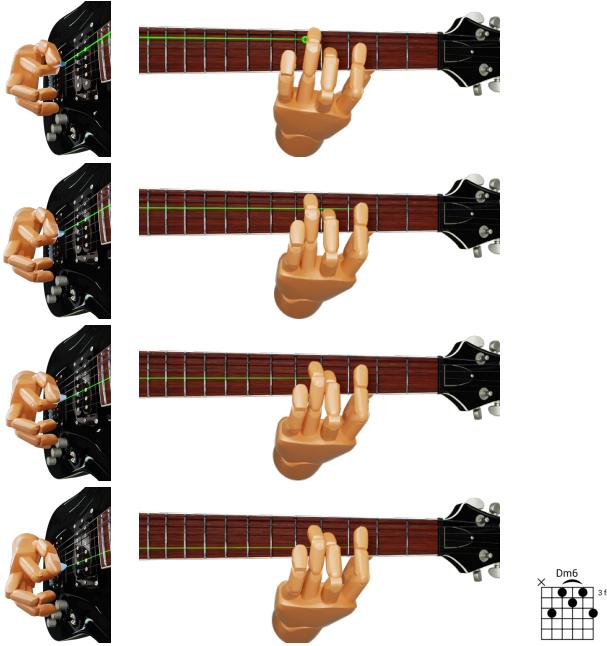


Fig. 13. Demonstrations of arpeggios. While the right hand picks individual strings following the target input notes (from top to bottom), the left hand keeps a stable pose for chord pressing by taking the chord as its target goal.

used for policy training. We also include some challenging chord-pressing tasks. For example, the 1st one in the 2nd row requires a long stretch between fingers; the 2nd one in the 1st row and the 3rd in the 2nd row require one finger to press multiple strings on the same fret; and the 1st one in the 4th row require different fingers to press strings on one fret (the ring finger on the 2nd and 3rd strings and pinkie on the 6th). In Fig. 11, we exhibit the results when the left-hand policy is trained without motion imitation. By comparing those results, we can see that our approach can effectively synthesize the limited poses in the reference data and generate human-like dexterous motions.

For the right hand, maintaining a steady rhythm during picking and strumming is a crucial skill for guitar playing. While Fig. 9 shows the picking performance of the right hand going up and down rhythmically, Fig. 10 shows the demonstrations of the right hand performing strumming where multiple strings are picked rapidly for successive notes. Without explicitly indicating picking directions, the right-hand policy can perform picking energy efficiently and maintain a regular up-to-down and down-to-up rhythm. In Fig. 13, we exhibit the policy’s performance of maintaining consistent rhythm while playing arpeggios, where the strings of one chord are picked individually by the right hand and the left hand keeps a chord-pressing pose. In this demonstration, we explicitly feed a *Dm6* chord (shown at the right-bottom corner in the figure) as the goal state input to the left-hand policy only, while the synchronizer and right-hand policy just take the picked string as the goal. Such

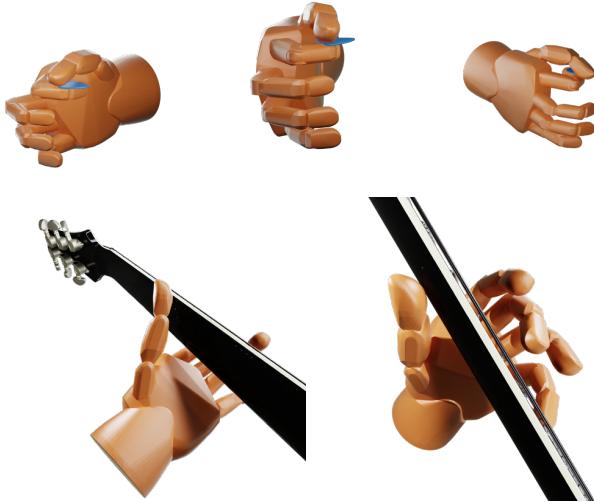


Fig. 14. Poses of the right hand holding the pick stably (top) and poses of the left hand contacting the guitar fretboard in a natural way.

an implementation is the same as what we could do in real life to put a chord indicator along the tab as guidance for fretting.

In Fig. 12, we compare the policy’s performance before and after synchronization. The highly synchronized behaviors with exceptional precision in timing enable our approach for music ensembles using multiple guitars. We refer to the supplementary video for the animated results of ensemble playing.

## 7 CONCLUSIONS

We present a novel approach for physics-based dual hand control in tasks requiring cooperation between two hands. Our approach first decouples the control of two hands in decentralized environments and then synchronizes the pre-trained single-hand policies’ behaviors for cooperation in a centralized environment. We demonstrate the proficiency of our approach in challenging tasks of guitar playing that require dexterous bimanual control with high precision both temporally and spatially. We refer to the supplementary materials for ablation studies and discussions about the limitations of the proposed method. Overall, our approach allows effective motion synthesis from unstructured references, providing high-quality guitar-playing motions in terms of both accuracy and naturalness for novel songs with diverse genres and tempos. Additionally, we captured a set of diverse guitar-playing motions from a professional guitarist. All the data is released publicly available for future work.

## ACKNOWLEDGMENTS

This work was supported in part by Stanford Institute for Human-Centered Artificial Intelligence. The motion capture solutions and devices were provided by NOKOV<sup>1</sup>. The authors would like to thank Ryan Canales from Clemson University for proposing the initial idea to perform guitar playing and for providing the initial data to verify the proposed method, and Prof. Ioannis Karamouzas from UC

<sup>1</sup><https://www.nokov.com>

Riverside and Prof. Sophie Jöerg from University of Bamberg for the meaningful discussions. We would also like to thank Hao Wang, the subject guitarist of our motion capture, for his performance, Mengyao Zhang and Dr. Hongming Fan from UIUC for their help in organizing motion capturing, and Dr. Shuchun Sun and Jichao Zhao from Clemson University for their help in developing the motion capture pipeline.

## REFERENCES

- OpenAI: Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, et al. 2020. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research* 39, 1 (2020), 3–20.
- Yu-Wei Chao, Wei Yang, Yu Xiang, Pavlo Molchanov, Ankur Handa, Jonathan Tremblay, Yashraj S Narang, Karl Van Wyk, Umar Iqbal, Stan Birchfield, et al. 2021. DexYCB: A benchmark for capturing hand grasping of objects. In *Conference on Computer Vision and Pattern Recognition*.
- Jiali Chen, Changjie Fan, Zhimeng Zhang, Gongzheng Li, Zeng Zhao, Zhigang Deng, and Yu Ding. 2023a. A Music-Driven Deep Generative Adversarial Model for Guzheng Playing Animation. *IEEE Transactions on Visualization and Computer Graphics* 29, 2 (2023), 1400–1414.
- Sirui Chen, Albert Wu, and C Karen Liu. 2023b. Synthesizing Dexterous Nonprehensile Pregrasp for Ungraspable Objects. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–10.
- Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. In *NIPS 2014 Workshop on Deep Learning*.
- George ElKoura and Karan Singh. 2003. Handrix: Animating the Human Hand. (2003), 110–119.
- Zicong Fan, Omid Taheri, Dimitrios Tzionas, Muhammed Kocabas, Manuel Kaufmann, Michael J Black, and Otmar Hilliges. 2023. ARCTIC: A dataset for dexterous bimanual hand-object manipulation. In *Conference on Computer Vision and Pattern Recognition*. 12943–12954.
- Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, Vijay Baiyya, Siddhant Bansal, Bikram Booté, Eugene Byrne, Zachary Chavis, Joya Chen, Feng Cheng, Fu-Jen Chu, Sean Crane, Avijit Dasgupta, Jing Dong, María Escobar, Cristhian Forigua, Abrahm Kahsay Gebrselasie, Sanjay Haresh, Jing Huang, Md Mohaiminul Islam, Suyog Dutt Jain, Rawal Khirodkar, Devansh Kukreja, Kevin J Liang, Jia-Wei Liu, Sagnik Majumder, Yongsen Mao, Miguel Martin, Effrosyni Mavroudi, Tushar Nagarajan, Francesco Ragusa, Santhosh K. Ramakrishnan, Luigi Seminara, Arjun Somayazulu, Yale Song, Shan Su, Zihui Xue, Edward Zhang, Jinxu Zhang, Angela Castillo, Changan Chen, Xinzhu Fu, Ryosuke Furuta, Cristina Gonzalez, Prince Gupta, Jiabo Hu, Yifei Huang, Yiming Huang, Weslie Khoo, Anush Kumar, Robert Kuo, Sach Lakhavani, Miao Liu, Mingjing Luo, Zhengyi Luo, Brighid Meredith, Austin Miller, Oluwatumini Oguntola, Xiaqing Pan, Penny Peng, Shraman Pramanick, Merey Ramazanova, Fiona Ryan, Wei Shan, Kiran Somasundaram, Chenan Song, Audrey Southerland, Masatoshi Tateno, Huiyu Wang, Yuchen Wang, Takuma Yagi, Mingfei Yan, Xitong Yang, Zecheng Yu, Shengxin Cindy Zha, Chen Zhao, Ziwei Zhao, Zifan Zhu, Jeff Zhuo, Pablo Arbeláez, Gedas Bertasius, David J. Crandall, Dima Damen, Jakob Julian Engel, Giovanni Maria Farinella, Antonino Furnari, Bernard Ghanem, Judy Hoffman, C. V. Jawahar, Richard A. Newcombe, Hyun Soo Park, James M. Rehg, Yoichi Sato, Manolis Savva, Jianbo Shi, Mike Zheng Shou, and Michael Wray. 2024. Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives. In *Conference on Computer Vision and Pattern Recognition*.
- Hank Heijink and Ruid GJ Meulenbroek. 2002. On the complexity of classical guitar playing: functional adaptations to task constraints. *Journal of motor behavior* 34, 4 (2002), 339–351.
- Asuka Hirata, Keitaro Tanaka, Masatoshi Hamanaka, and Shigeo Morishima. 2022. Audio-Driven Violin Performance Animation with Clear Fingering and Bowing. In *ACM SIGGRAPH 2022 Posters*. 1–2.
- Hsuan-Kai Kao and Li Su. 2020. Temporally guided music-to-body-movement generation. In *Proceedings of the 28th ACM International Conference on Multimedia*. 147–155.
- Diederik P Kingma. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- Paul G Kry and Dinesh K Pai. 2006. Interaction capture and synthesis. *ACM Transactions on Graphics (TOG)* 25, 3 (2006), 872–880.
- Taras Kucherenko, Patrik Jonell, Sanne Van Waveren, Gustav Eje Henter, Simon Alexandersson, Iolanda Leite, and Hedvig Kjellström. 2020. Gesticulator: A framework for semantically-aware speech-driven gesture generation. In *Proceedings of the 2020 international conference on multimodal interaction*. 242–250.

- Vikash Kumar and Emanuel Todorov. 2015. Mujoco haptix: A virtual reality system for hand manipulation. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 657–663.
- Bochen Li, Akira Maezawa, and Zhiyao Duan. 2018. Skeleton Plays Piano: Online Generation of Pianist Body Movements from MIDI Performance.. In *ISMIR*. 218–224.
- Yuanzhi Liang, Qianyu Feng, Linchao Zhu, Li Hu, Pan Pan, and Yi Yang. 2022. Seeg: Semantic energized co-speech gesture generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10473–10482.
- Hung Yu Ling, Fabio Zinno, George Cheng, and Michiel van de Panne. 2020. Character controllers using motion VAEs. *ACM Trans. Graph.* 39, 4, Article 40 (2020).
- C Karen Liu. 2008. Synthesis of interactive hand manipulation. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 163–171.
- C Karen Liu. 2009. Dexterous manipulation from a grasping pose. In *ACM SIGGRAPH 2009 papers*. 1–6.
- Jun-Wei Liu, Hung-Yi Lin, Yu-Fen Huang, Hsuan-Kai Kao, and Li Su. 2020. Body Movement Generation for Expressive Violin Performance Applying Neural Networks. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2020), 3787–3791.
- Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. 2021. Isaac Gym: High Performance GPU-Based Physics Simulation For Robot Learning. *arXiv:2108.10470 [cs.RO]*
- Josh Merel, Yuval Tassa, Dhruva TB, Sriram Srinivasan, Jay Lemmon, Ziyu Wang, Greg Wayne, and Nicolas Heess. 2017. Learning human behaviors from motion capture by adversarial imitation. *arXiv preprint arXiv:1707.02201* (2017).
- Igor Mordatch, Zoran Popović, and Emanuel Todorov. 2012. Contact-invariant optimization for hand manipulation. In *Proceedings of the ACM SIGGRAPH/Eurographics symposium on computer animation*. 137–144.
- Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. 2018. DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills. *ACM Trans. Graph.* 37, 4, Article 143 (2018).
- Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. 2022. ASE: Large-Scale Reusable Adversarial Skill Embeddings for Physically Simulated Characters. *ACM Trans. Graph.* 41, 4, Article 94 (2022).
- Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. 2021. AMP: Adversarial Motion Priors for Stylized Physics-Based Character Control. *ACM Trans. Graph.* 40, 4, Article 144 (2021).
- Nancy S Pollard and Victor Brian Zordan. 2005. Physically based grasping control from example. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*. 311–318.
- Shenhan Qian, Zhi Tu, Yihao Zhi, Wen Liu, and Shenghua Gao. 2021. Speech drives templates: Co-speech gesture synthesis with learned templates. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 11077–11086.
- Aleksander Radisavljevic and Peter F Driessens. 2004. Path difference learning for guitar fingering problem. In *International Conference on Mathematics and Computing*.
- Aravind Rajeswaran, Sarvjeet Ghora, Balaraman Ravindran, and Sergey Levine. 2017. EPOpt: Learning Robust Neural Network Policies Using Model Ensembles. In *Int. Conf. on Learning Representations*.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. *arXiv:1707.06347 [cs.LG]*
- Eli Shlizerman, Lucio Dery, Hayden Schoen, and Ira Kemelmacher-Shlizerman. 2018. Audio to body dynamics. In *Conference on Computer Vision and Pattern Recognition*. 7574–7583.
- Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. 2017. Hand keypoint detection in single images using multiview bootstrapping. In *Conference on Computer Vision and Pattern Recognition*.
- Lucchas Ribeiro Skreinig, Denis Kalkofen, Ana Stanescu, Peter Mohr, Frank Heyen, Shohei Mori, Michael Sedlmair, Dieter Schmalstieg, and Alexander Plopski. 2023. guitarARhero: Interactive Augmented Reality Guitar Tutorials. 29, 11 (2023), 4676–4685.
- Omid Taheri, Nima Ghorbani, Michael J Black, and Dimitrios Tzionas. 2020. GRAB: A dataset of whole-body human grasping of objects. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV* 16. Springer, 581–600.
- Yangang Wang, Jianyuan Min, Jianjie Zhang, Yebin Liu, Feng Xu, Qionghai Dai, and Jinxiang Chai. 2013. Video-based hand manipulation capture through composite motion control. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 1–14.
- Jungdam Won, Deepak Gopinath, and Jessica Hodgins. 2020. A scalable approach to control diverse behaviors for physically simulated characters. *ACM Transactions on Graphics* 39, 4 (2020), 33–1.
- Kevin A. Wortman and Nicholas Smith. 2021. CombinoChord: A Guitar Chord Generator App. *IEEE 11th Annual Computing and Communication Workshop and Conference* (2021), 0785–0789.
- Zhaoming Xie, Jonathan Tseng, Sebastian Starke, Michiel van de Panne, and C Karen Liu. 2023. Hierarchical planning and control for box loco-manipulation. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 6, 3 (2023), 1–18.
- Pei Xu and Ioannis Karamouzas. 2021. A GAN-Like Approach for Physics-Based Imitation Learning and Interactive Character Control. *Proc. of the ACM on Computer Graphics and Interactive Techniques* 4, 3 (2021).
- Pei Xu, Xiumin Shang, Victor Zordan, and Ioannis Karamouzas. 2023a. Composite Motion Learning with Task Control. *ACM Transactions on Graphics* 42, 4 (2023).
- Pei Xu, Kaixiang Xie, Sheldon Andrews, Paul G. Kry, Michael Neff, Ioannis Karamouzas, and Victor Zordan. 2023b. AdaptNet: Policy Adaptation for Physics-Based Character Control. *ACM Transactions on Graphics* 42, 6 (2023).
- Zeshi Yang, Kangkang Yin, and Libin Liu. 2022. Learning to use chopsticks in diverse gripping styles. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–17.
- Heyuan Yao, Zhenhua Song, Baoquan Chen, and Libin Liu. 2022. ControlVAE: Model-Based Learning of Generative Controllers for Physics-Based Characters. *ACM Trans. Graph.* 41, 6, Article 183 (2022).
- Heyuan Yao, Zhenhua Song, Yuyang Zhou, Tenglong Ao, Baoquan Chen, and Libin Liu. 2023. MoConVQ: Unified Physics-Based Motion Control via Scalable Discrete Representations. *arXiv preprint arXiv:2310.10198* (2023).
- Yuting Ye and C Karen Liu. 2012. Synthesis of detailed hand manipulations using contact sampling. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 1–10.
- Kevin Zakka, Philipp Wu, Laura Smith, Nimrod Gileadi, Taylor Howell, Xue Bin Peng, Sumeet Singh, Yuval Tassa, Pete Florence, Andy Zeng, et al. 2023. Robopianist: Dexterous piano playing with deep reinforcement learning. In *Conference on Robot Learning*.
- H Zhang, Y Ye, T Shiratori, and T Komura. 2021. ManipNet: neural manipulation synthesis with a hand-object spatial representation. *ACM Transactions on Graphics* (2021).
- Zeyi Zhang, Tenglong Ao, Yuyao Zhang, Qingzhe Gao, Chuan Lin, Baoquan Chen, and Libin Liu. 2024. Semantic Gesticulator: Semantics-Aware Co-Speech Gesture Synthesis. *ACM Trans. Graph.* (2024), 17 pages. <https://doi.org/10.1145/3658134>
- Wenping Zhao, Jianjie Zhang, Jianyuan Min, and Jinxiang Chai. 2013. Robust realtime physics-based motion control for human grasping. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 1–12.
- Yuanfeng Zhu, Ajay Sundar Ramakrishnan, Bernd Hamann, and Michael Neff. 2013. A system for automatic animation of piano performances. *Computer Animation and Virtual Worlds* 24, 5 (2013), 445–457.

## A GUITAR PLAYING MOTION DATASET

Our motion data were captured from a professional guitarist. We took the motion capture solutions provided by NOKOV and employed 18 infrared cameras for motion capture, 7 of which have a resolution of 9MP and 11 have a resolution of 4MP. We placed 19 markers on each hand of the subject, and 4 markers on the guitar for localization. All data were captured at 120FPS. The camera deployment and marker placement are exhibited in Fig. S1. We collected diverse general guitar-playing motions about 1 hour long, including:

- 12 major scales,
- chromatic scales,
- diverse chords,
- arpeggios,
- strumming and picking,
- bends,
- sliding,
- vibrato,
- palm mute,
- natural harmonics,
- artificial harmonics,
- hammer-ons and pull-offs.

All notes were played as 1/8th and mostly at 100BPM. The instrument is a *Fender American Deluxe Stratocaster<sup>2</sup>* guitar with 6 strings and 22 frets. The scale length is 25.5 inches.

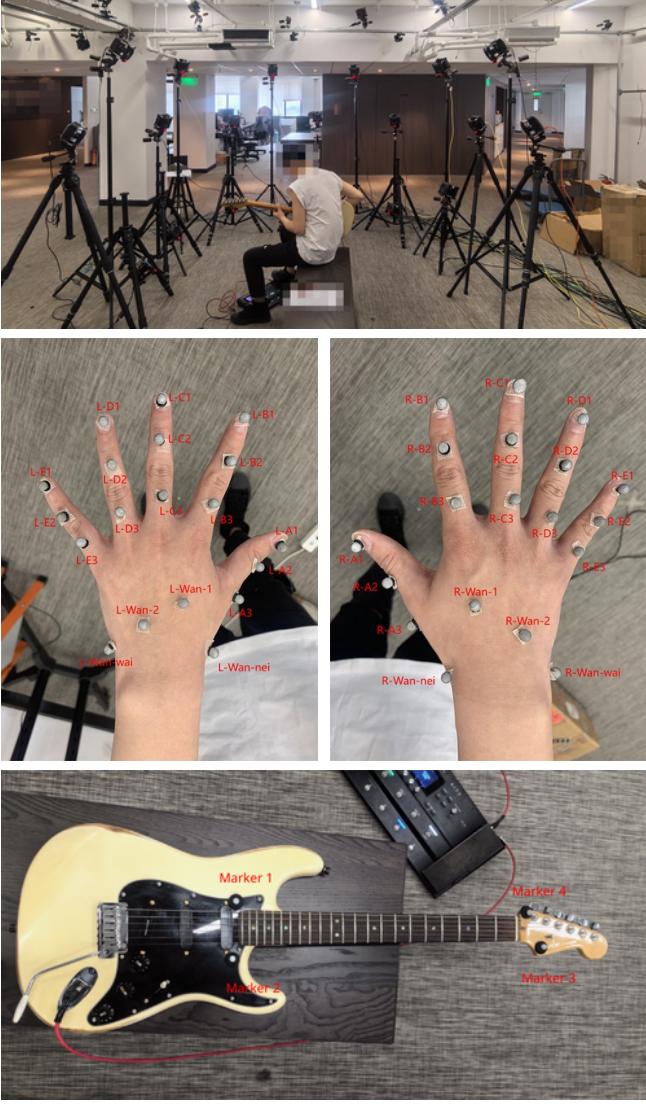


Fig. S1. Setup of our motion capture pipeline. Top: camera deployment. Middle: marker placement on hands. Bottom: marker placement on guitar.

## B IMPLEMENTATION DETAILS

All our control policies are optimized using PPO [Schulman et al. 2017] as the backbone reinforcement learning algorithm with network parameters updated by Adam optimizers [Kingma 2014]. The hyperparameters used for policy training are listed in Table S1. Half of the training samples used by the discriminator are drawn from the simulated character and half are sampled from the reference motions. The network structures are shown in Fig. S2. We use a GAN-like architecture combining reinforcement learning and motion imitation for policy training [Xu and Karamouzas 2021]. We use the same network structure for the left and right-hand policies, while our policy synchronization scheme does not require them to be

<sup>2</sup><https://serialnumberlookup.fender.com/product/0101000748>

Table S1. Hyperparameters

Parameter	Value
policy network learning rate	$5 \times 10^{-6}$
critic network learning rate	$1 \times 10^{-4}$
discriminator learning rate	$1 \times 10^{-5}$
reward discount factor ( $\gamma$ )	0.95
GAE discount factor ( $\lambda$ )	0.95
surrogate clip range ( $\epsilon$ )	0.2
gradient penalty coefficient ( $\lambda^{GP}$ )	10
number of PPO workers (simulation instances)	512
PPO replay buffer size	4096
PPO batch size	256
PPO optimization epochs	5
discriminator replay buffer size	8192
discriminator batch size	512

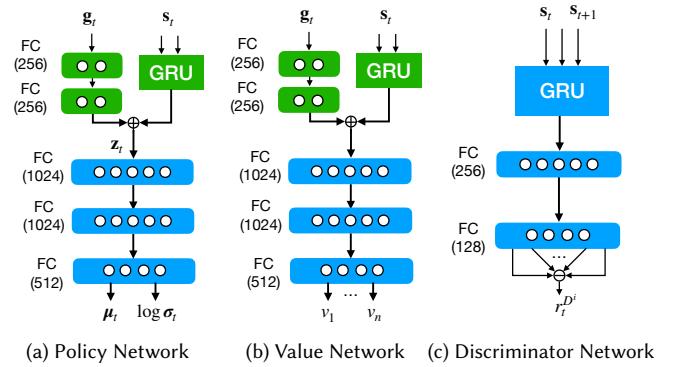


Fig. S2. Network structures. We use  $\oplus$  denoting the add operator and  $\ominus$  denoting the average operator. The state encoder consists of blocks in green, where a GRU encoder is employed for pose state encoding temporally, and a two-layer MLP encoder is employed for goal state encoding.

identical. The state encoder of each policy consists of a GRU [Chung et al. 2014] encoder with a 256-dimensional hidden state for pose state encoding and an MLP encoder for goal state encoding. The two encoded states are added together as the latent for manipulation during policy synchronization. The value networks have the same architecture as the policy networks, but have multiple outputs depending on the number of objectives (8 for the left-hand policy, 2 for the right-hand policy, and 9 for the joint policy synchronized using pre-trained single-hand policies). Especially, when training two-hand policies, it takes  $s_t = \{s_t^L, s_t^R\}$  and  $g_t = \{g_t^L, g_t^R\}$  as input. We normalize the fretting representation of  $g_t^L$  into the range of  $[-1, 1]$  for the left-hand policy, while the right-hand policy uses a binary representation of  $g_t^R$  to indicate if a string should be picked or not. We run a moving-average normalizer on  $s_t^L$  and  $s_t^R$  during policy training. For value networks, we perform output normalization using the technique of PopArt [Rajeswaran et al. 2017].

All policies were trained on a machine equipped with a Nvidia V100 GPU. It takes around 3–4 days to perform individual policy training, consuming about  $8 \times 10^8$  samples. The synchronization

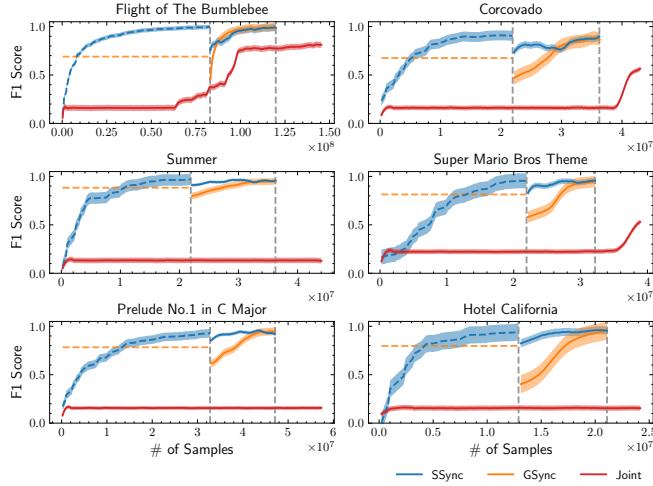


Fig. S3. Training performance of the two-hand policies using the pre-trained general-purpose left-hand policy for synchronization (*GSync*) and that using policies trained only for specific songs (*SSync*). The dashed orange and blue lines indicate the corresponding left-hand policies' performance when evaluated solely without the right hand. For the pre-trained general-purpose policy, we show its evaluation performance using dashed orange lines. For the specifically trained policies, we show their training performance using dashed blue lines. The gray lines are an indicator to distinguish the single-hand policy training phase and the synchronization phase. Red lines are the learning performance when two hands are put together as a joint agent trained from scratch.

can be done quickly, typically within 1.5 to 12 hours consuming about  $1 \times 10^7$  to  $8 \times 10^7$  samples, depending on the difficulty of the songs (mainly related to the chord patterns and tempos).

### C ABLATION STUDIES

Our proposed policy synchronization approach can synchronize the behaviors of two hands for cooperative bimanual control, and, meanwhile, fine-tune the pre-trained single-hand policies for specific songs. Both of them would bring about an improvement in the F1 scores of the achieved two-hand policies. To show the performance of policy synchronization solely, we study the synchronization performance using left-hand policies that are trained specially for specific songs. We still use the right-hand policy that was pre-trained for general purposes, as it already can perform string picking accurately for all the tested songs.

In Fig. S3, we compare the two-hand performance during training using the pre-trained general-purpose left-hand policy (*GSync*) and that using the ones trained for specific songs (*SSync*). The performances of the corresponding single-hand policies are shown in dashed lines. As can be seen, typically, training a left-hand policy for a specific song (dashed lines in blue) would consume around  $2 \times 10^7$  samples. For difficult pieces, e.g. the very fast song *Flight of The Bumblebee* that has a tempo of 170BPM and the classic music *Prelude No.1 in C Major* that has lots of arpeggios, it would need around  $3.5 \times 10^7$  and even  $8 \times 10^7$  samples. The specially trained policies can reach an F1 score above 0.98 or even equal to 1, which is

nearly perfect in terms of string-pressing accuracy. However, when put together with the right hand for cooperation, they still suffer a significant performance drop of around 20%, because of the desynchronization between the two hands' behaviors. The performance of the pre-trained general-purpose left-hand policy (dashed lines in orange) is worse than the specially trained left-hand policies, especially for songs with many chords, e.g. *Corcovado*, but still can provide an F1 score above 0.8 for most of the tested songs. While being able to reach a similar final performance with the specially trained left-hand policies through synchronization, the general-purpose policy is reusable for different songs. We can also see that the worse performance before policy synchronization, as shown in Fig. 6 mainly comes from the unsynchronized behaviors of the single-hand policies. Our policy synchronization scheme can effectively coordinate the behaviors of two hands within a small budget of training samples, and achieve high F1 scores at the end.

Additionally, in Fig. S3, we show the training curves when two hands are directly trained together from scratch as a joint agent (*Joint*). In that case, we employ the same network structure as that used for single-hand policy training, but use a concatenated pose state of  $s_t^L$  and  $s_t^K$ . The policy is optimized under the same training scheme depicted in Section 5.3 but has one more imitation objective for the right hand. As shown in the figure, training a two-hand policy directly from scratch is difficult. Though we can see a performance improvement for some tested songs as the training goes on, the curve increases very slowly. The two hands, at the initial training phase, mainly perform random exploration for basic skill learning rather than trying to build effective cooperation, and thus the learning is inefficient within the limited budget of training samples. Note that simulating scenarios with two hands is around 40% slower than that with only one hand. This means that when the same number of samples are consumed, it would take much more time to train a two-hand policy. Therefore, even if the *Joint* policy may finally achieve better performance as the training goes on, our proposed approach is more efficient by reducing the training needed for the joint policy training of two hands in the centralized environment. We refer to the supplementary video for visual comparison.

### D LIMITATIONS

For the left hand, policies trained by our approach can perform fret pressing in natural human poses with high accuracy. The resulting motions, though dexterous, are not always able to fully reflect the grace of human guitarists in terms of smooth hand movement. Ideally, accurate fretting requires swift movement of the hand and fingers to press target notes responsively, minimizing the transitioning time between consecutive notes as much as possible. This, however, will result in fast and abrupt movement. While professional human guitarists would exhibit more sophisticated planning behaviors for energy-efficient motion control, our energy-related reward term (Eq. 8) simply penalizes the hand and fingers' fast movement at the cost of responsiveness. The associated weight for the term has to be small enough to guarantee the policy's responsiveness even when facing notes lasting for short durations (cf. Eq. 7 and 15, or has to be fine-tuned carefully for specific rhythms to balance responsiveness and energy efficiency. Due to this trade-off, our control policies

cannot completely avoid unnatural fast or abrupt movement of the hand and fingers, especially for fast songs, as we can see from the animated results shown in the supplementary video. Besides, our policies only have a limited horizon of future notes (5 in our implementation), while experienced human guitarists would perform more careful and intelligent motion planning for fretting based on a longer horizon and other factors like finger dexterity and muscle fatigue. In the arpeggio demo shown in Fig. 13, due to the limited observation horizon, the left-hand policy itself is unable to detect chord patterns based on the target notes of individual strings. We have to specially feed a chord note as the goal input for the left hand while the right hand takes the normal goal notes of individual strings, or the left hand will press single strings without keeping a consistent chord-pressing pose. Though this is analogous to explicitly putting a chord symbol on the tab to guide fretting in real-world practice, it would be an interesting direction for future work to further introduce human-level intelligence for motion planning during guitar playing.

Another limitation of this work is the lack of physically simulated string dynamics. This makes us have to rely on heuristic rules to decide the pressed and picked states of strings and generate sound by post-processing, rather than directly getting their physical vibration states for accurate sound generation. The lack of string

dynamics constrains us from introducing guitar-playing techniques that involve complex interactions with strings, like vibrato and fingerstyle. Meanwhile, we use rigid bodies instead of deformable ones for hand modeling, which may overlook subtle changes in hand shape. Besides, all simulations run discretely with a fixed FPS (60Hz in our implementation), limiting the control's time resolution. For example, given music where the shortest notes are 1/16th notes, tempos between 100 and 105BPM will be rounded as 100 to ensure that the shortest notes last for an integer number of frames. It is a great direction for future work to address the fine control problem with arbitrary temporal resolution requirements.

As a general approach for two-agent cooperative learning, our study in the guitar-playing task demonstrates that the proposed policy synchronization scheme can efficiently adapt two single-agent policies into a joint policy for cooperative tasks. However, this scheme requires decentralized environments for individual policy training, which are not always easy to set up. As a bimanual control method for instrument playing, our approach is expected to work for instruments similar to guitar, like bass and ukulele, but may not be generalized to those involving complex string dynamics like harp and violin. Developing unified policies for various instruments is an interesting direction for future research.