A Robotic System For Imitating Human Percussionists

Fouad Sukkar[1,2], Richard Savery[3], Nadimul Haque[1,2], Cedric Le Gentil[1], Raphael Falque[1] and Teresa Vidal-Calleja[1,2]
Robotics Institute[1] and Australian Cobotics Centre[2], University of Technology Sydney
MCCALL, Macquarie University[3]
{fouad.sukkar@,akmnadimul.haque@student.,cedric.legentil@,raphael.guenot-falque@, teresa.vidalcalleja@}uts.edu.au, richard.savery@mq.edu.au

Abstract

Robot musicians have the potential to revolutionise the way humans perceive and create music. Recent breakthroughs in this field have tended to focus more on the digital generation of music. Instead, we address how a musician's physical embodiment can be translated to a robotic arm. Robots endowed with human-like musical capability open the possibility for wider applications such as human-robot bands, musical education and musical art. Prior work in this area tends to rely on pre-programmed actuation which is limited to simple motion and sound. In this paper, we propose a robotic system capable of imitating a human musician, with a focus on percussion instruments. Our system consists of a method for recording the human demonstration, a compact continuous representation of the demonstrated motion and a motion reproduction method which considers the dynamic constraints of the robot. We present results of our system and show that it is capable of closely reproducing the motion of the human percussionist.

1 Introduction

Robotic musicianship is an emerging field where robots manifest music in much the same way humans have throughout history. Robotic musicians have a wide potential of applications, including collaborating with musicians during live performances (human-robot bands), musical education and musical art. However, recent breakthroughs in this area have been oriented towards digital generation of music [Briot, 2021], such as through generative artificial intelligence (AI) [Shahriar, 2022], instead of through physical actuation. This leaves much to be desired, especially when highly capable robots are becoming increasingly widely available.

In this paper, we propose a robotic system capable of imitating human musicians, with a focus on percussion instruments. Prior work typically focuses on either machine musicianship (machines with musical skills such as listening, performance and composition) or musical mechatronics (physical devices that generate sound). For example, existing robot musicians typically use pre-programmed actuation and thus are limited to simple motion and sounds. Our work aims to bridge machine musicianship and musical mechatronics by manifesting highly capable cobots with human-level musical skills.

Figure 1: UR3 Robot arm imitating a human percussionist playing a caxixi (percussion instrument).

Endowing robots with skills through human demonstration is a well-studied field in robotics [Ravichandar *et al.*, 2020]. However, most methods tend to be sensitive to imperfect demonstrations as a result of noisy perception. Furthermore, our problem setting is challenging because there is no direct mapping between the human motion and the robot, known as the correspondence problem [Nehaniv *et al.*, 2002]. This is further compounded by the fact that the robot may have kinematic constraints, such as fewer degrees of freedom (DoF), and dynamic constraints, such as velocity and acceleration limits. By leveraging recent work in this area we aim to design a system which facilitates overcoming these challenges.

Our system is inspired by a hybrid reinforcement learning (RL) and learning from demonstration (LfD) framework [Nematollahi *et al.*, 2022] which performs online refinements to the demonstrated motion via feedback from physical interactions with the environment. In the context of this work, sound discrepancy between the human and the robot could be used as feedback. While a refinement process is not implemented in this paper, we propose a system for enabling such a process and reserve addressing this scenario in future work.

In [Nematollahi *et al.*, 2022], authors utilise a Gaussian Mixture Model (GMM) for a compact representation of the human trajectory. However, we show that such a representation is unsuitable for reproducing the complex motions of the human percussionist. Instead, we propose to use a Gaussian Process (GP) with optimised inducing points which has far better representational capabilities.

In this paper, we describe the proposed system, including our method for recording the motion of a human percussionist, a suitable trajectory representation for use within an RL refinement framework and a motion reproduction method which respects the dynamic constraints of the robot. To the best of our knowledge, this work is the first attempt at programming a robotic musician directly from human demonstration. We demonstrate the efficacy of our method in producing a good baseline motion reproduction of the human percussionist.

2 Related Work

In this section we cover related work in the relevant areas of robots imitating humans and robotic musicians. First we discuss trajectory representations and learning from human demonstration. Then we turn our attention to machine musicianship and musical mechatronics.

2.1 Trajectory Representation

Dynamical system-based control methods leverage statistical models as a compact and robust trajectory representation. Several parameterised models have been utilised in prior work, including GMMs [Li *et al.*, 2021; Nematollahi *et al.*, 2022], GPs [Mukadam *et al.*, 2016; Bhardwaj *et al.*, 2020] and dynamic movement primitives (DMP) [Su *et al.*, 2020]. While the latter two provide a more accurate representation, DMPs can be less robust to perturbations [Ginesi *et al.*, 2021]. GMMs tend to be lightweight; however, have lower representational power [Ratiu and Prichici, 2017]. GPs have better representational power; however, at the cost of increased computational complexity [Liu *et al.*, 2020a]. This complexity depends on the number of required inducing points (representative points in the trajectory) to effectively fit to the trajectory. In [Le Gentil and Vidal-Calleja, 2023], authors demonstrated that GPs can achieve high performance with a relatively small number of optimised inducing points. GPs additionally have other useful applications in robotics such as imposing velocity constraints and principled modelling of noise.

2.2 Learning from human demonstration

Learning from demonstration (LfD) is a popular approach for teaching robots skills through expert demonstrations. Commonly, LfD methods use kinaesthetic teaching or teleoperation to demonstrate the skill directly on the robot system [Rozo *et al.*, 2013; Zhu *et al.*, 2018; Savarimuthu *et al.*, 2018; Gao *et al.*, 2019; Liu *et al.*, 2020b; Su *et al.*, 2021]. While this avoids correspondence issues [Rakita *et al.*, 2017], by directly mapping the demonstration trajectory to the robot trajectory, it can be restrictive on the demonstrator, resulting in unnatural and impeded motion.

In the context of teaching a robot to imitate a human percussionist, this can be severely limiting. Instead we would like to enable musicians to play their instrument naturally during the demonstration. However, as a consequence several challenges arise, such as no direct mapping between the human and robot due to differences in kinematics and dynamics.

The work in [Sukkar *et al.*, 2023] partly addresses this issue; however, only accounts for kinematic discrepancies. Recent work proposed a hybrid RL and LfD approach which performs online refinements to the modelled human trajectory through physical interactions with the environment [Nematollahi *et al.*, 2022]. Such a method is promising for addressing this problem and correcting for discrepencies, for example due to sensor noise and model error, between the human and reproduced demonstration.

2.3 Machine Musicianship and Musical Mechatronics

Research at the intersection of music and robotics generally focuses on either machine musicianship or musical mechatronics. Machine musicianship emphasizes the development of robotics focusing on music perception, composition, performance, and theory [Savery *et al.*, 2021]. Musical mechatronics, instead addresses the construction of physical devices that generate sound through mechanical means [Kapur, 2005]. Machine musicianship and musical mechatronics are not necessarily binary, however research groups have tended to focus on one area.

Figure 2: Caxixi percussion instrument. Sounded by shaking the instrument.

In machine musicianship recent work has led to advancements such as a rapping robot, able to interact in real-time [Savery *et al.*, 2020]. A number of notable musical mechatronics efforts have addressed wind instruments [Dannenberg *et al.*, 2005; Solis *et al.*, 2010], and string instruments [Singer *et al.*, 2004; Kusuda, 2008].

The majority of developments in robotic percussionists have used solenoid actuation [Maes *et al.*, 2011; Kapur *et al.*, 2011; Singer *et al.*, 2004]. More recently brushless direct current (BLDC) motors have been used, and shown to allow for higher levels of musical expressiveness through an increased range of volumes and timbrel variation [Yang *et al.*, 2020]. Despite progress in this area, the above applications have been limited to simplistic pre-programmed motions. We aim to enable highly capable robotic musicians by learning directly from human demonstrations.

3 Problem Setup

In this section we describe the problem setup including a formal statement of the human percussionist imitation problem we are addressing. We additionally provide necessary background knowledge on GPs in order to describe our trajectory representation approach.

3.1 Problem Statement

We wish to utilise a robotic manipulator for mimicking a human playing a percussion instrument. The particular percussion instrument used in this paper is called a caxixi and is pictured in Fig. 2; however, our approach is percussion instrument agnostic. A caxixi is an indirectly struck idiophone consisting of a basket and a hard bottom with seeds inside. It is sounded by shaking the instrument and depending on the angle can produce a variety of sounds.

Given a captured human demonstration, the goal of the method is to play it back with the robot. Thus the demonstration trajectory needs to be recorded in a way that facilitates this playback. Furthermore, due to localisation errors and differences in dynamics and kinematic structure between the robot arm and the human arm, a direct trajectory mapping is not possible. Thus, there is a need to refine this trajectory such that the sound of the caxixi being played by the robot matches that being played by the percussionist.

In this paper, we do not address this refinement step. However, we aim to facilitate such a process through an appropriate parameterised trajectory representation, similar to the approach in [Nematollahi *et al.*, 2022]. Such a representation could be used in conjunction with a reinforcement learning (RL) algorithm which indirectly transforms the trajectory by perturbing the parameters, which tends to be a significantly smaller search space than the searching over the entire trajectory space. Thus, a trajectory representation with minimal parameters is desirable. Given a loss function, such as sound similarity [Slaney *et al.*, 2008], the RL algorithm could in turn learn to fine-tune the robot's motion.

Furthermore, the robot arm has joint velocity and acceleration limits. Thus we must ensure that the trajectory being reproduced respects these limits. To facilitate such a process we aim to design a system consisting of: a method for capturing the sound and motion of a percussionist, a suitable parameterised trajectory representation, and a motion reproduction method which respects the dynamic constraints of the robot whilst remaining as close to the demonstration.

3.2 Gaussian Process preliminaries

Gaussian Process regression is a non-parametric, probabilistic interpolation method [Rasmussen and Williams, 2006]. This subsection briefly provides background knowledge about GP regression using a simple 1D example. Let us consider a signal h as a function of time t. By modelling $h(t)$ with a GP as

$$h \sim \mathcal{GP}(0, k(t, t')), \tag{1}$$

the covariance kernel function $k(t, t')$ represents the covariance between two instances of the signal

$$\mathrm{cov}(h(t), h(t')) = k(t, t').$$

The definition of a GP states that any finite set of instances of the signal corresponds to a multivariate Gaussian distribution. Therefore, given N noisy measurements of the signal

$$y_i = h(t_i) + \eta, \quad \text{with} \quad \eta \sim \mathcal{N}(0, \sigma_y) \text{ and } i = 1, \cdots, N,$$

it is possible to express the noisy instances as well as a novel query point $h^*(t)$ as a multivariate Gaussian distribution

$$\begin{bmatrix} \mathbf{y} \\ h^*(t) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{K_{tt}} + \sigma_y^2 \mathbf{I} & (\mathbf{k_t}(t))^\top \\ \mathbf{k_t}(t) & k(t,t) \end{bmatrix} \right), \tag{2}$$

with $\mathbf{y} = [y_1, \cdots, y_N]^\top$ the vector of noisy observations, $\mathbf{k_t}(t) = [k(t_0, t), \cdots, k(t_N, t)]$ the covariance vector between any time t and the timestamps of the observations, and $\mathbf{K_{tt}} = [(\mathbf{k_t}(t_0))^\top, \cdots, (\mathbf{k_t}(t_0))^\top]$ the covariance matrix of the observations. The inference of $h^*(t)$ using GP regression corresponds to the conditioning of (2) with respect to the observations. Accordingly the mean and variance of $h^*(t)$ are derived as

$$\begin{aligned} h^*(t) &= \mathbf{k_t}(t)[\mathbf{K_{tt}} + \sigma_y^2\mathbf{I}]^{-1}\mathbf{y} \\ \mathrm{var}(h^*(t)) &= k(t,t) - \mathbf{k_t}(t)[\mathbf{K_{tt}} + \sigma_y^2\mathbf{I}]^{-1}(\mathbf{k_t}(t))^\top. \end{aligned} \tag{3}$$

These last equations allow for the inference of the signal h at any timestamp t in a data-driven manner without relying on any explicit model of the signal.

4 Robotic Percussionist System

Here we describe our system for imitating a human percussionist. First we describe our method for recording the human demonstration. Then we describe our compact and smooth trajectory representation method. Lastly, we describe our motion reproduction method which considers the dynamic contraints of the robot.

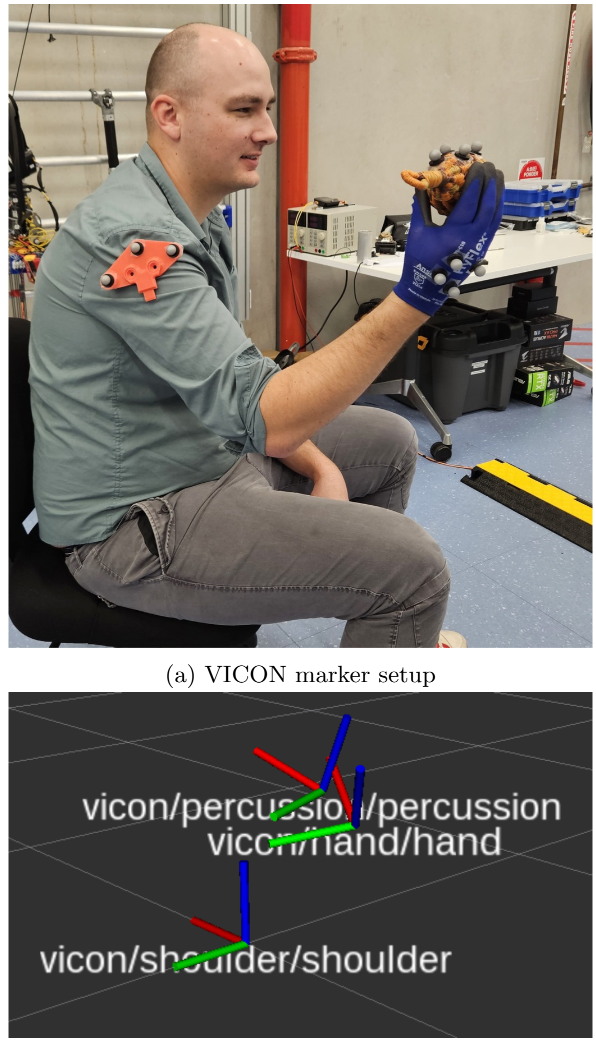
4.1 Recording Human Percussionist

For capturing the demonstrated percussion piece we track the trajectories of both the percussionist's back palm and instrument. We do so using a VICON system and by placing markers on the percussionist's arm, as shown in Fig. 3. Trajectories of the hand and instrument are tracked relative to the shoulder which maps to the base frame of the robot arm. The VICON collects 6-DoF poses at a frequency of 300 Hz and we consider demonstrations of around 20 s in duration. We denote the position and orientation measurements from the VICON system at time $\mathbf{t}_i$ with $\tilde{\mathbf{p}}_i$ and $\tilde{\mathbf{r}}_i$ $(i = 1, \cdots, Q)$, respectively.

4.2 Trajectory Model

In this work, we propose to use GPs to represent the motion of the human percussionist as a continuous 6-DoF trajectory due to their greater representational capacity over other similar methods. Similar to [Le Gentil and Vidal-Calleja, 2023], our method uses a sparse set of optimised inducing values to perform GP regression instead of using the raw pose measurements. This results in a lower dimensional representation of the trajectory which is crucial in reducing the search space for a refinement RL algorithm whose actions would perturb these inducing points.

In this work, the inducing points correspond to pairs of arbitrarily-set timestamps and optimised values. Formally, let us model the trajectory with 6 independent



(a) VICON marker setup

(b) TF frames in Rviz

Figure 3: VICON marker setup for recording human demonstration and corresponding rigid body frames shown in Rviz.

GPs

$$\begin{aligned} p_\bullet &\sim \mathcal{GP}(0, k(t, t')) \\ r_\bullet &\sim \mathcal{GP}(0, k(t, t')), \end{aligned}$$

where $\mathbf{p}(t) = [p_x, p_y, p_z]^\top$ is the instrument position, and $\mathbf{r}(t) = [r_x, r_y, r_z]^\top$ the orientation using the rotation vector formalism. By defining a set of inducing points at fixed timestamps t_i with $t_i - t_{i-1} = \Delta t$[1], the position and orientation of the instrument through time are defined following (3)

$$\begin{aligned} p_\bullet^*(t) &= \mathbf{k_t}(t)[\mathbf{K_{tt}} + \sigma_y^2\mathbf{I}]^{-1}\boldsymbol{\gamma}_{p_\bullet} \\ r_\bullet^*(t) &= \mathbf{k_t}(t)[\mathbf{K_{tt}} + \sigma_y^2\mathbf{I}]^{-1}\boldsymbol{\gamma}_{r_\bullet}, \end{aligned}$$

[1] $\Delta t = 0.1$ s in our experiments.

with $\gamma_{p_\bullet}$ and $\gamma_{r_\bullet}$ the vectors of position and orientation inducing values. In order to accurately represent the VICON measurements of the instrument's trajectory with our GP models, the inducing values are optimised with non-linear optimisations:

$$\{\gamma^*_{p_x}, \gamma^*_{p_y}, \gamma^*_{p_z}\} = \underset{\{\gamma_{p_x}, \gamma_{p_y}, \gamma_{p_z}\}}{\operatorname{argmin}} \left(\sum_{i=1}^{Q} \|\mathbf{p}^*_i(\mathrm{t}_i) - \tilde{\mathbf{p}}_i\|^2 \right)$$

$$\{\gamma^*_{r_x}, \gamma^*_{r_y}, \gamma^*_{r_z}\} = \underset{\{\gamma_{r_x}, \gamma_{r_y}, \gamma_{r_z}\}}{\operatorname{argmin}} \left(\sum_{i=1}^{Q} \|\mathbf{r}^*_i(\mathrm{t}_i) - \tilde{\mathbf{r}}_i\|^2 \right).$$

Note that to ensure the continuity of the rotation representation (avoiding the 2π wrapping) we initialise the inducing values similarly to [Le Gentil and Vidal-Calleja, 2023].

The computational complexity of GP regression is cubic, $\mathcal{O}(N^3)$, in the number of inducing points due to the matrix inversion present in (3). While the sparsity of the inducing points helps reduce this computation, more importantly the times of the inducing points can remain fixed during the fine-tuning procedure. Thus this inversion would only need to be performed once for the entire learning phase.

4.3 Motion Reproduction

Given a fitted GP model to the human demonstration, we can reproduce a trajectory at any arbitrary time using (3). Although the GP produces inherently smooth trajectories, it is still possible that intermediate waypoints can violate the joint velocity limits of the robot.

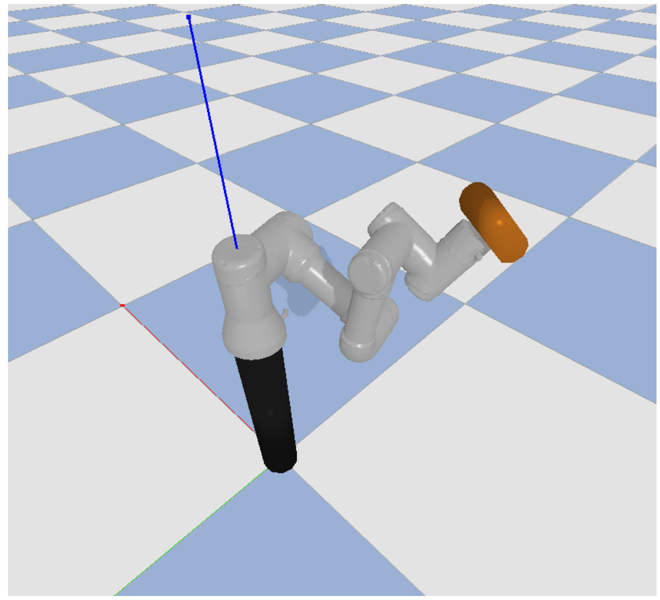
Thus, we first compute the corresponding joint space trajectory of the robot using an inverse kinematics solver. This joint space trajectory is then passed through a moving average filter with an experimentally chosen averaging window. As a final processing step, any remaining waypoints that violate the robot's velocity constraints are removed from the trajectory.

5 Experiments

In this section we validate our proposed trajectory model. First, we describe the proposed experimental setup. Then, we present and discuss results demonstrating the suitability of our GP-based approach compared to a GMM-based model.

5.1 Hardware and Software Setup

The proposed hardware setup for reproducing motion, shown in Fig. 4, is composed of: a Universal Robots UR3 robot arm, caxixi mounted on UR3 end-effector and microphone (although not used at this stage). For producing collision-free trajectories we utilise PyBullet with OpenAI Gym and IKFast plugin [Cambel, 2023]. For sending trajectories to the robot we use the Universal Robots ROS driver [UniversalRobots, 2023].



(a) Simulated UR3 robot in PyBullet

(b) UR3, caxixi and Microphone

Figure 4: Experimental setup

5.2 Results

To evaluate our approach we recorded three percussion demonstrations of approximately 20 second duration each. We used 200 inducing points for our GP model. For comparison we fit a GMM with the same number of Gaussian components. The root-mean-square error for each model's predicted position and orientation is shown in Table 1. The GP greatly outperforms the GMM with consistently low error.

We show the fit of our proposed GP model compared to the GMM on an example recorded demonstration trajectory in Fig. 5. As can be seen, the GP is able to follow the measured trajectory very closely whilst the GMM struggles and over-regularises the trajectory. However,

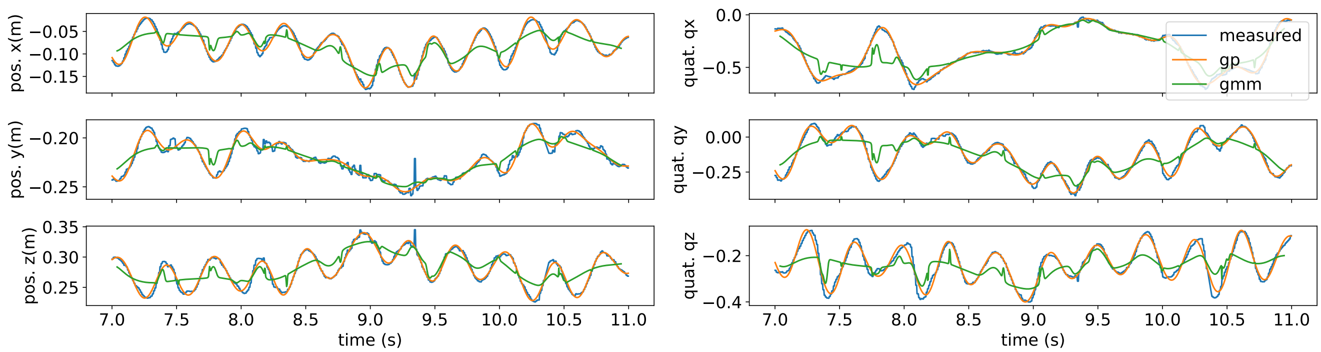


Figure 5: GMM versus GP model fit to a portion of an example demonstrated motion (measured) with same number of parameters.

Table 1: Quantitative comparison of GP and GMM trajectory reproduction. Values are RMSE between the raw measurements from the demonstration and values queried from the models at the same timestamps.

Recording	GP Position (m)	GMM Position (m)	GP Orientation (rad)	GMM Orientation (rad)
1	**0.006**	0.0384	**0.056**	0.340
2	**0.006**	0.083	**0.053**	0.793
3	**0.007**	0.034	**0.0522**	0.304

it should be noted that this does not necessarily translate to accurate sound reproduction, which is ultimately what we desire. Furthermore, the smoothing of the GP, whilst effectively filters out noise, does result in deviation from some parts of the trajectory.

Using a $\Delta T = 0.1\,\mathrm{s}$, we are able to accurately represent the trajectory with only approximately 3.5% of the number of raw poses which is a large reduction in the search space for the refinement process. In contrast, the GMM, utilising the same number of parameters, results in poor motion reproduction.

With our GP implementation, the inference of 6000 6-DoF poses with 200 inducing points takes only approximately 80 ms. This is negligible with respect to the trajectory duration which were in the order of 20 s.

6 Conclusion and Future Work

In this paper we proposed a robotic system capable of closely imitating the motion of a human percussionist. In our experimental results we showed that our proposed trajectory representation model was effective in reproducing the motion with a relatively low number of parameters. This is crucial for future work where we aim to implement a refinement process, via physical interaction, which not only corrects for the percussionist's motion but also the sound produced by the instrument.

We additionally aim to take a more principled approach to dynamic constraint satisfaction in the motion reproduction. A possible approach is to apply linear operators [Särkkä, 2011] to the kernel function of the GP and optimise for the given limits in joint velocity space. Lastly, we believe our method could be applied to a wider range of percussion instruments and potentially other musical instruments.

Acknowledgments

This work was partially funded by the ARC Industrial Transformation Training Centre for Collaborative Robotics in Advanced Manufacturing under grant IC200100001. Cedric Le Gentil is supported by the Australian Research Council Discovery Project under Grant DP210101336.

References

[Bhardwaj *et al.*, 2020] Mohak Bhardwaj, Byron Boots, and Mustafa Mukadam. Differentiable gaussian process motion planning. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 10598–10604. IEEE, 2020.

[Briot, 2021] Jean-Pierre Briot. From artificial neural networks to deep learning for music generation: history, concepts and trends. *Neural Computing and Applications*, 33(1):39–65, 2021.

[Cambel, 2023] Cambel. Cambel/ur_ikfast: Python ikfast library for universal robots. `https://github.com/cambel/ur_ikfast`, 2023.

[Dannenberg *et al.*, 2005] Roger B Dannenberg, Ben Brown, Garth Zeglin, and Ron Lupish. Mcblare: a robotic bagpipe player. In *Proceedings of the 2005 conference on New interfaces for musical expression*, pages 80–84. National University of Singapore, 2005.

[Gao *et al.*, 2019] Xiao Gao, Jie Ling, Xiaohui Xiao, and Miao Li. Learning Force-Relevant Skills from Human Demonstration. *Complexity*, 2019, 2019.

[Ginesi *et al.*, 2021] Michele Ginesi, Nicola Sansonetto, and Paolo Fiorini. Overcoming some drawbacks of dynamic movement primitives. *Robotics and Autonomous Systems*, 144:103844, 2021.

[Kapur *et al.*, 2011] Ajay Kapur, Michael Darling, Jim W Murphy, Jordan Hochenbaum, Dimitri Diakopoulos, and T Trimpin. The karmetik notomoton: A new breed of musical robot for teaching and performance. In *NIME*, pages 228–231, 2011.

[Kapur, 2005] Ajay Kapur. A history of robotic musical instruments. In *ICMC*. Citeseer, 2005.

[Kusuda, 2008] Yoshihiro Kusuda. Toyota's violin-playing robot. *Industrial Robot: An International Journal*, 35(6):504–506, 2008.

[Le Gentil and Vidal-Calleja, 2023] Cedric Le Gentil and Teresa Vidal-Calleja. Continuous latent state preintegration for inertial-aided systems. *The International Journal of Robotics Research*, 0(0):02783649231199537, 2023.

[Li *et al.*, 2021] Jiehao Li, Junzheng Wang, Shoukun Wang, and Chenguang Yang. Human–robot skill transmission for mobile robot via learning by demonstration. *Neural Computing and Applications*, pages 1–11, 2021.

[Liu *et al.*, 2020a] Haitao Liu, Yew-Soon Ong, Xiaobo Shen, and Jianfei Cai. When gaussian process meets big data: A review of scalable gps. *IEEE transactions on neural networks and learning systems*, 31(11):4405–4423, 2020.

[Liu *et al.*, 2020b] Nailong Liu, Xiaodong Zhou, Zhaoming Liu, Hongwei Wang, and Long Cui. Learning peg-in-hole assembly using Cartesian DMPs with feedback mechanism. *Assembly Automation*, 6(September):895–904, 2020.

[Maes *et al.*, 2011] Laura Maes, Godfried-Willem Raes, and Troy Rogers. The man and machine robot orchestra at logos. *Computer Music Journal*, 35(4):28–48, 2011.

[Mukadam *et al.*, 2016] Mustafa Mukadam, Xinyan Yan, and Byron Boots. Gaussian process motion planning. In *2016 IEEE international conference on robotics and automation (ICRA)*, pages 9–15. IEEE, 2016.

[Nehaniv *et al.*, 2002] Chrystopher L Nehaniv, Kerstin Dautenhahn, et al. The correspondence problem. *Imitation in animals and artifacts*, 41, 2002.

[Nematollahi *et al.*, 2022] Iman Nematollahi, Erick Rosete-Beas, Adrian Röfer, Tim Welschehold, Abhinav Valada, and Wolfram Burgard. Robot skill adaptation via soft actor-critic gaussian mixture models. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8651–8657. IEEE, 2022.

[Rakita *et al.*, 2017] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. A motion retargeting method for effective mimicry-based teleoperation of robot arms. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 361–370, 2017.

[Rasmussen and Williams, 2006] C E Rasmussen and C K I Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[Ratiu and Prichici, 2017] Mariana Ratiu and Mariana Adriana Prichici. Industrial robot trajectory optimization-a review. In *MATEC web of conferences*, volume 126, page 02005. EDP Sciences, 2017.

[Ravichandar *et al.*, 2020] Harish Ravichandar, Athanasios S Polydoros, Sonia Chernova, and Aude Billard. Recent advances in robot learning from demonstration. *Annual review of control, robotics, and autonomous systems*, 3:297–330, 2020.

[Rozo *et al.*, 2013] Leonel Rozo, Pablo Jiménez, and Carme Torras. A robot learning from demonstration framework to perform force-based manipulation tasks. *Intelligent Service Robotics*, 6(1):33–51, 2013.

[Särkkä, 2011] Simo Särkkä. Linear operators and stochastic partial differential equations in Gaussian process regression. *Artificial Neural Networks and Machine Learning–ICANN 2011*, pages 151–158, 2011.

[Savarimuthu *et al.*, 2018] Thiusius Rajeeth Savarimuthu, Anders Glent Buch, Christian Schlette, Nils Wantia, Jürgen Robmann, David Martínez, Guillem Alenyà, Carme Torras, Ales Ude, Bojan Nemec, Aljaž Kramberger, Florentin Wörgötter, Eren Erdal Aksoy, Jeremie Papon, Simon Haller, Justus Piater, and Norbert Krüger. Teaching a Robot the Semantics of Assembly Tasks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(5):670–692, 2018.

[Savery *et al.*, 2020] Richard Savery, Lisa Zahray, and Gil Weinberg. Shimon the rapper: a real-time system for human-robot interactive rap battles. In *International Conference on Computational Creativity (11th: 2020)*, pages 212–219. University of Coimbra, 2020.

[Savery *et al.*, 2021] Richard Savery, Lisa Zahray, and Gil Weinberg. *Shimon Sings-Robotic Musicianship Finds Its Voice*, pages 823–847. Springer International Publishing, Cham, 2021.

[Shahriar, 2022] Sakib Shahriar. Gan computers generate arts? a survey on visual arts, music, and literary text generation using generative adversarial network. *Displays*, 73:102237, 2022.

[Singer *et al.*, 2004] Eric Singer, Jeff Feddersen, Chad Redmon, and Bil Bowen. Lemur's musical robots. In *NIME*, pages 181–184, 2004.

[Slaney *et al.*, 2008] Malcolm Slaney, Kilian Weinberger, and William White. Learning a metric for music similarity. In *International Symposium on Music Information Retrieval (ISMIR)*, volume 148, 2008.

[Solis *et al.*, 2010] Jorge Solis, Takeshi Ninomiya, Klaus Petersen, Maasaki Takeuchi, and Atsuo Takanishi. Development of the anthropomorphic saxophonist robot was-1: Mechanical design of the simulated organs and implementation of air pressure feedback control. *Advanced Robotics*, 24(5-6):629–650, 2010.

[Su *et al.*, 2020] Hang Su, Yingbai Hu, Zhijun Li, Alois Knoll, Giancarlo Ferrigno, and Elena De Momi. Reinforcement learning based manipulation skill transferring for robot-assisted minimally invasive surgery. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2203–2208. IEEE, 2020.

[Su *et al.*, 2021] Hang Su, Andrea Mariani, Salih Ertug Ovur, Arianna Menciassi, Giancarlo Ferrigno, and Elena De Momi. Toward Teaching by Demonstration for Robot-Assisted Minimally Invasive Surgery. *IEEE Transactions on Automation Science and Engineering*, 18(2):484–494, 2021.

[Sukkar *et al.*, 2023] Fouad Sukkar, Victor Hernandez Moreno, Teresa Vidal-Calleja, and Jochen Deuse. Guided learning from demonstration for robust transferability. In *2023 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2023.

[UniversalRobots, 2023] UniversalRobots. Universalrobots/universal_robots_ros_driver. `https://github.com/UniversalRobots/Universal_Robots_ROS_Driver`, 2023.

[Yang *et al.*, 2020] Ning Yang, Richard Savery, Raghavasimhan Sankaranarayanan, Lisa Zahray, and Gil Weinberg. Mechatronics-driven musical expressivity for robotic percussionists. In *NIME*, 2020.

[Zhu *et al.*, 2018] Alex Zihao Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception. *IEEE Robotics and Automation Letters*, 3(3):2032–2039, 2018.