ACHIEVING HUMAN LEVEL COMPETITIVE ROBOT TABLE TENNIS

David B. D'Ambrosio[1,*], **Saminda Abeyruwan**[1,*], **Laura Graesser**[1,*], **Atil Iscen**[1], **Heni Ben Amor**[2], **Alex Bewley**[2], **Barney J. Reed**[2,†], **Krista Reymann**[2], **Leila Takayama**[2,§], **Yuval Tassa**[2], **Krzysztof Choromanski**, **Erwin Coumans**, **Deepali Jain**, **Navdeep Jaitly**, **Natasha Jaques**, **Satoshi Kataoka**, **Yuheng Kuang**, **Nevena Lazic**, **Reza Mahjourian**, **Sherry Moore**, **Kenneth Oslund**, **Anish Shankar**, **Vikas Sindhwani**, **Vincent Vanhoucke**, **Grace Vesom**, **Peng Xu**, **Pannag R. Sanketi**[1]

Google DeepMind
[1]Primary contributors, [*]Corresponding authors (order randomized, equal contributions),
[2]Core contributors (Alphabetized),
[†]Work done at Google DeepMind via Stickman Skills Center LLC,
[§]Work done at Google DeepMind via Hoku Labs.

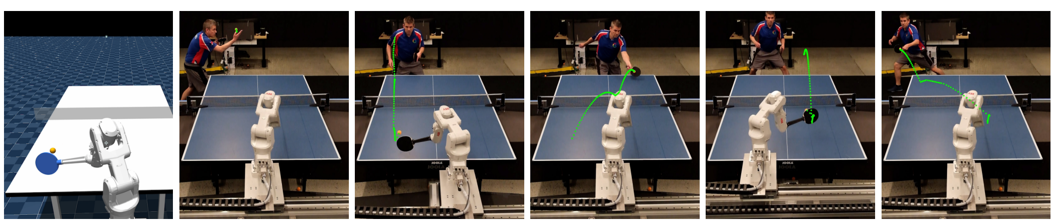

Figure 1: The robot playing in simulation (left) and against a professional coach in real. Green dots show the ball trajectory.

ABSTRACT

Achieving human-level performance on real world tasks is a north star for the robotics community. We present the first learned robot agent that reaches amateur human-level performance in competitive table tennis. Table tennis is a physically demanding sport that requires humans years to master. We contribute (1) a hierarchical and modular policy architecture consisting of (i) low level controllers with their skill descriptors that model their capabilities and (ii) a high level controller that chooses the low level skills, (2) techniques for enabling zero-shot sim-to-real and curriculum building, including an iterative approach (train in sim, deploy in real), and (3) real time adaptation to unseen opponents. Policy performance was assessed through 29 robot vs. human matches of which the robot won 45% (13/29). All humans were unseen players and their skill level varied from beginner to tournament level. Whilst the robot lost all matches vs. the most advanced players it won 100% matches vs. beginners and 55% matches vs. intermediate players, demonstrating solidly amateur human-level performance.

1 INTRODUCTION

Robot learning has made inspiring progress, yet achieving human-level performance in complex domains, like table tennis, which demand high-speed motion, precise control, and human-robot interaction, remains challenging Fu et al. (2024); Wu et al. (2023); Li et al. (2023). Table tennis has served as a valuable benchmark for robotics research since the 1980s, with numerous robots developed to tackle various aspects of the game Billingsley (1983); Huang et al. (2015); Ding et al. (2022); Chen et al. (2021); Abeyruwan et al. (2023b); D'Ambrosio et al. (2023). However, *no prior work has addressed playing a full competitive game against a previously unseen human opponent.*

This paper presents the first learned robot agent capable of playing competitive table tennis at a human level (see Figure 1). Control architectures and hierarchies play a critical role in robotics Brooks (1986); Arkin (1998). We introduce a hierarchical and modular policy architecture, consisting of multiple low-level skill policies and a high-level controller which chooses the best skill to execute, to address the challenge of combining strategic decision-making with physical skills execution, and the challenge of learning a model for a diverse skill set. Training is efficient, since all low-level policies start from a small set of base models and are then specialized. Our work draws inspiration from previous research on hierarchical robot policies, which traditionally relied on engineered policies and arbitration modules Brooks (1986); Arkin (1998); Rosenblatt & Thorpe (1997); Kawato et al. (1987); Daniel et al. (2016); Ahn et al. (2022). Our architecture is closest in spirit to the work in Mülling et al. (2013) in which a gating network is learned to create mixtures of existing low-level policies. The gating network generates probabilities indicating the likelihood that a policy is the right one given the current context. We utilize instance-based learning and tree-search over skill descriptors for low-level policies, enabling real-time learning and adaptation, distinguishing our approach from prior methodologies.

A hybrid training method is employed, *synergistically combining reinforcement learning in simulation and deployment in real*. The robot's skills are iteratively refined in simulation based on real-world data, creating an automatic task curriculum and enabling continuous improvement. This enables efficient training and zero-shot transfer to real hardware Sutton & Barto (2018); Osa et al. (2018); Caluwaerts et al. (2023); Kumar et al. (2021); Cheng et al. (2023); Stepputtis et al. (2020); Collaboration et al. (2023); Zhao et al. (2020); Sontakke et al. (2023); Jiang et al. (2021); Peng et al. (2018); Lee et al. (2018); D'Ambrosio et al. (2023); Todorov et al. (2012); Abeyruwan et al. (2023b). We improve upon previous iterative approaches by utilizing seed human vs. human play data, zero-shot policy transfer, and a non-parametric dataset-based ball distribution for better alignment between simulation and real-world human play.

Furthermore, the robot *adapts to unseen human opponents* by tracking match statistics and estimating preferences online, facilitating real-time learning and adaptation to the environment and opponent Kumar et al. (2021). This capability is crucial for playing games in the physical world with humans, a challenging problem with ongoing research in cooperative games and ad-hoc team-play Stone et al. (2010); Hu et al. (2020); Carroll et al. (2019); Strouse et al. (2021).

We build upon previous research in table tennis robotics that covers various aspects such as action and motion generation, state estimation, and human strategy identification Andersson (1988); Hashimoto et al. (1987); Knight & Lowery (1986); Schweitzer & Wen (1994); Mülling et al. (2013); Tebbe et al. (2018); Büchler et al. (2022); D'Ambrosio et al. (2023); Muelling et al. (2010); Huang et al. (2015); Koç et al. (2018); Zhu et al. (2018); Tebbe et al. (2021); Ding et al. (2022); Liu et al. (2013); Abeyruwan et al. (2023b); Matsushima et al. (2003); Sun et al. (2011); Chen et al. (2021); Büchler et al. (2022); Nakashima et al. (2011); Gao et al. (2019); Blank et al. (2017); Gossard et al. (2024); Muelling et al. (2014); Wang et al. (2013; 2017); Guist et al. (2024). To date, the Omron Forpheus robot Kyohei et al. (2019); Liu et al. (2013) has the closest capabilities to the agent presented in this work, demonstrating sustained rallies. A key point of difference is that our agent learns the control policies and perception system, whereas the Forpheus agent uses a model-based approach. Also, our agent is able to play full matches. Our work also contributes to the growing body of sports research tackling complex, dynamic tasks involving human interaction. While most work focuses on sub-aspects or simplified settings, we aim to achieve *competitive gameplay against humans in realistic conditions*, similar to the RoboCup competition and other robot sports Kitano et al. (1997); Röfer et al. (2023); Stone et al. (2005); Behnke et al. (2006); Suriani et al. (2024);

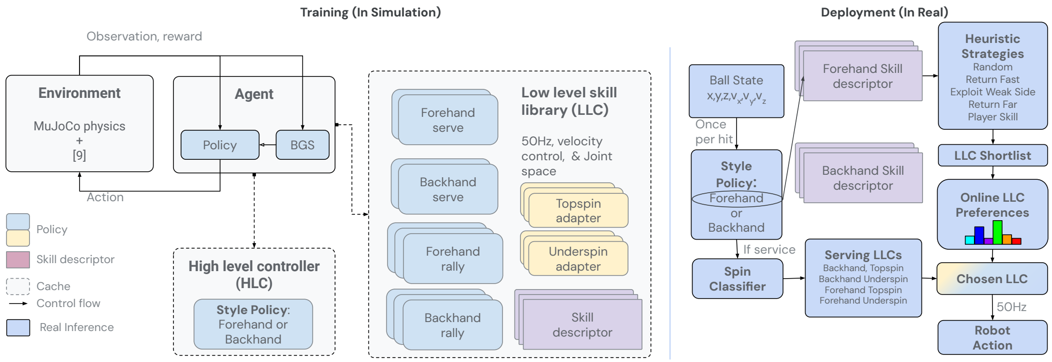

Figure 2: Method overview. We train a skill library of low-level controllers (LLCs), including serving and rallying, and sim-to-sim adapters from a dataset of ball states. Using the same ball states, we train a high level controller (HLC) for style selection. The policies are trained in simulation and transferred zero-shot to the physical world. When deployed, the HLC decides which LLC should return the ball by first applying a style policy to the current ball state to determine forehand or backhand. If the ball is a service, it will classify the spin and pick the corresponding LLC. Otherwise it must determine which of the many rallying LLCs will perform best by finding the most similar ball state within the corresponding set of LLC skill tables and getting the return statistics. Heuristic strategies are applied to these statistics to produce a shortlist of candidate LLCs. The final LLC is chosen based on preferences per LLC learned online.

Wang et al. (2024); Haarnoja et al. (2024); Yang et al. (2021; 2022); Petric et al. (2012); Zaidi et al. (2023); Abeyruwan et al. (2023a); Mori et al. (2019); Kaufmann et al. (2023).

In summary, this paper introduces a novel robot learning system that achieves amateur human-level performance in competitive table tennis against unseen opponents. We make four main technical contributions; (1) a hierarchical and modular policy architecture, (2) techniques for zero-shot sim-to-real transfer and automatic curriculum, (3) real-time adaptation to opponents, and (4) a user study to evaluate the system's performance and engagement.

2 METHOD

2.1 HARDWARE AND ENVIRONMENT

Figure 1 depicts the physical robot. The table tennis robot is a 6 DoF ABB IRB 1100 arm mounted on top of two Festo linear gantries, enabling motion in the 2d plane. The x gantry, which moves side to side across the table, is 4m long and the y gantry, which moves towards and away from the table, is 2m long. A 3D printed paddle handle and paddle with short pips rubber Glo (2024) is attached to the ABB arm. A pair of Ximea MQ013CG-ON cameras operating at 125Hz capture images of the ball and these are used as input into a neural-perception system which produces ball positions at the same frequency. We use a PhaseSpace motion capture system consisting of 20 cameras mounted around the play area to track the human opponent's paddle.

We model table tennis as a single-agent sequential decision making problem in which the human opponent is modeled as part of the environment using the *Markov Decision Process* (MDP) Puterman (2014) formalization. In what follows we describe changes to this system that were made to enable real-time competitive play with humans.

2.2 LLC TRAINING

The table tennis agent shown in Figure 2 consists of two levels of control which we refer to as the high level controller (HLC) and the low level controllers (LLCs). LLCs provide a library of skills that our HLC can choose from.

Training algorithm and LLC architecture All LLCs were trained in simulation with Blackbox Gradient Sensing (BGS) Abeyruwan et al. (2023b), an evolutionary strategies (ES) algorithm, on the task described in Section 2.1. The training task distribution of initial ball states is sampled from a real world dataset, gathered iteratively through multiple cycles of policy training and real world evaluations (discussed in Section 2.4). BGS was chosen because it produced policies with relatively smooth actions and has been shown to have strong sim-to-real transfer performance D'Ambrosio et al. (2023). Each policy is a 1D dilated-gated CNN Oord et al. (2016) with 10k parameters following Gao et al. (2020) plus an optional FILM adapter layer of 2.8k parameters to aid sim-to-real transfer. The observation space is $(8, 16)$ consisting of 8 consecutive timesteps of ball position and velocity (6), robot joint position (8), and one-hot style; forehand or backhand (2). The action space is $(8,)$ representing joint velocities. All policies are run at 50Hz.

Training generalist base LLCs First we trained generalist base LLCs for each style (forehand, backhand). To train for a particular style, each ball state in the dataset was annotated with *forehand*, *backhand*, or *center* based on where the ball trajectory intersected with the back of the table on the robot side. Forehand LLCs were trained on only *forehand* + *center* balls, backhand on *backhand* + *center*. This created an overlap in the center where policies of either style are capable of returning the same balls. The policy was also rewarded for moving towards a reference pose (either forehand or backhand) at the beginning of the shot. These base LLCs are important, not only to have a strong starting polices capable of returning a wide range of balls to branch from, but also to anchor play in specific styles for efficient returns.

Training specialists Next we specialized LLCs to different skills by adding reward function components and / or adjusting the training data mix and fine-tuning a new policy initialized from one of the existing LLCs. We experimented with the types of skills to train for based on advice from a table tennis coach and general game intuition, including targeting specific return locations, maximizing return velocity, and specializing to return serves of either topspin or underspin, fast balls, and lobs. We found we did not need a specialist for lobs, and were unable to train a specialist on fast balls due to lack of data and hardware limitations. We therefore focused on developing serving, targeting and fast hitting specialists in addition to the generalists.

Determining the total set of skill policies The final system contained 17 LLCs. 4 were specialized for returning serves, 13 for rallying. 11 played with a forehand style, 6 with a backhand style. Importantly, each policy had the same initial robot pose, enabling straightforward sequencing of LLC choices, since the initial robot pose will be in-distribution for all LLCs. We kept training LLCs until we had covered our target set of skills. Due the modular architecture, there was little downside in including additional LLCs. If we had a strong LLC, we included it, even if there was already an LLC covering that particular skill.

2.3 THE HIGH LEVEL CONTROLLER (HLC)

The HLC is responsible for making strategic decisions. Concretely, the HLC is responsible for selecting which LLC should be run for each incoming ball. The HLC does not have a fixed control frequency but instead is triggered to act, once, every time the opponent hits the ball. Within the HLC, there are six components that are combined to produce the choice of LLC — style policy, spin classifier, LLC skill descriptors, match statistics, strategies, LLC preferences (H-values). Figure 2 (RHS) presents an overview of the control flow depicting how each of these elements is combined.

Style policy The style policy determines if the robot should return the ball with a forehand or backhand style. The architecture is similar to the LLCs but with only 4.5k parameters and has a (8, 128) observation space. We flatten the LLC (8, 16) observation (described in Section 2.2) and stack the latest 8 observations to form the observation. The action space is (2,) representing a one-hot categorical choice between forehand and backhand. To train this policy, we selected a general-purpose forehand and backhand LLC and froze their weights, then trained the style policy to maximize the expected ball landing rate using all available ball states (including reflections). We found the policy generalized to serving ball states so used a single style policy for both serving and rallying phases of the game.

Spin classifier The spin classifier is a binary classifier that determines if the incoming serve was hit by the humans as a topspin or an underspin. To train the model, we built a dataset of paddle and ball states from the serving dataset (see Section 2.4). Specifically, we record a history of the 6 timestamps of ball and paddle states directly before the paddle made contact with the ball. The observation space is (18,) the policy is a 2-layer MLP which outputs the probability that the incoming ball is topspin or underspin.

LLC skill descriptors To excel in interactive sports, it is crucial to understand one's own capabilities. This motivated the development of LLC skill descriptors which provide detailed metrics to the HLC on the estimated performance of each LLC for a given incoming ball. To create the descriptors, we evaluated each LLC in simulation on all 28k ball states averaged over ten repetitions, recording the following policy metadata — initial ball position and velocity, median hit velocity, ball landing location, ball landing rate. This metadata was used to construct lookup tables (we used KD-Trees Bentley (1975)) with keys representing initial ball position and velocity. Given any ball in play, the table can be queried for information about the likely performance of each LLC were it to be selected by averaging performance of similar balls it has seen in the past. We used n=1 nearest neighbors due to inference time constraints.

We observed a sim-to-real gap remained in LLC performance. LLC hit rates in the real world were high, however ball return rates, whilst good, were lower than the $> 80\%$ we typically observed in simulation. This meant that building skill descriptors using only simulated data was likely to lead to errors. To address this we updated each LLC's skill descriptor using real-world data. Four researchers played with the robot and gathered 91 - 257 real world ball throws per LLC. For each LLC and for each ball collected, the 25 nearest neighbors in the relevant LLC-specific tree were updated, weighting the simulated metrics and real world metrics for a single ball throw equally.

Strategies and LLC shortlist Every time the HLC acts, five hand-coded heuristics were used to generate a shortlist (one per heuristic) of the most promising LLC candidates, given the output from the style policy and information collected by the HLC about the opponent on their ability to return balls both in total and broken down by forehand, backhand, and center returns. This opponent information is persisted between games with the same opponent. Note, not all heuristics use all available information.

The set of heuristics we utilized are as follows — random selection, prioritization of hit velocity, prioritization of landing distance farthest from the initial ball state, exploitation of opponent's weak side, consideration of opponent's skill by selecting LLCs with the farthest landing position for the given ball state if the opponent's hit rate exceeds 75%.

From the shortlist we select the LLC that will be used to return the ball with weighted sampling (to make the robot less predictable) described below.

LLC preferences (H-value) & choosing an LLC Another key aspect of playing competitive sports is understanding the opponent's capabilities and being able to adapt in response. This motivated learning online preferences for each LLC which, as well as helping to bridge the remaining sim-to-real gap, provide a rudimentary model of the human opponent.

We learned a numerical preference using a simple gradient bandit algorithm Sutton & Barto (2018) for each LLC, $H(LLC) \in \mathbb{R}$, based on the LLC's online performance. The agent selects LLCs more often if their preference is higher.

For a given ball, each LLC in the shortlist is associated with an offline return rate. We combined the offline return rate and the online preferences (H-values) to select an LLC. We found that combining learned H-values with information from the skill descriptor tables played an important role in improving performance. These H-values serve two major purposes. (1) Online sim-to-real correction; even though efforts were made through the offline updates to the skill descriptor tables, a sim-to-real gap remained, likely because the sample of real world balls used to update the tables was small and generated by a small number of players. H-values allow the policy to quickly switch away from poor-performing LLCs to more stable ones. (2) To learn player-specific strengths and weaknesses; if the current opponent is able to easily send shots that one LLC struggles to return, the HLC can shift weight to another the opponent can less easily exploit.

Each time an LLC was selected the H-value was updated using the binary ball land signal as the reward function. For each new opponent, these values were initialized to a set of known baseline preferences, to ensure everyone played against the same initial agent. These preferences were updated and persisted across games for the same opponent.

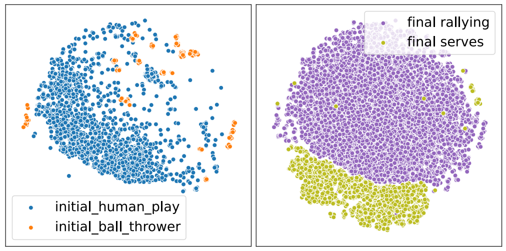

Figure 3: Visualization of the task distribution dataset. TSNE van der Maaten & Hinton (2008) was used to project from 9-D ball states to 2-D.

2.4 TECHNIQUES FOR ENABLING ZERO-SHOT SIM-TO-REAL

There are two core challenges in simulating robotic table tennis. First, faithfully modeling the robot, paddle, and ball dynamics. Second, accurately modeling the task distribution, i.e. the distribution over initial states of real-world incoming ball trajectories toward the robotic player.

Modeling ball and robot dynamics We recreated the simulation environment described in D'Ambrosio et al. (2023) and Abeyruwan et al. (2023b) within MuJoCo Todorov et al. (2012). This enhances the prior simulated work by leveraging a more advanced solid state fluid dynamics for ball trajectory simulation, refining model and system identification, and improving the representation of paddle rubber. System identification was performed for each actuator-joint pair following the methodology presented in Haarnoja et al. (2024).

The paddle rubber was explicitly modeled using two orthogonal passive joints representing a spring-damper system to approximate a rubber surface. Ball-rubber contact solver parameters (softness, slip, friction) were determined empirically, while joint stiffness, damping, and armature were established through parameter sweeps optimizing for sim-to-real transfer. Analogously, ball-table contact solver parameters were also measured. We observed a bimodal distribution in contact solver parameters for the paddle rubber restitution when we completed system ID for topspin and underspin ball contacts. Underspin balls exhibit a damping coefficient of -103, while topspin balls have a damping coefficient of approximately -0. Consequently, during the ***topspin correction*** phase of policy training (described below), the simulator dynamically selects the appropriate solver parameters based on the ball's pre-contact spin. This bimodality was not observed in the ball-table contact solver parameters.

We utilized domain randomization, observation noise, and latency similar to D'Ambrosio et al. (2023). We randomized table and paddle damping, and friction parameters during training. We employed two shaping rewards, net height reward and a target for the last ABB joint at ball-paddle contact, to mitigate a sim-to-real gap observed due to robot returns overshooting the opponent's side. This approach not only addressed the intended criteria but also promoted competitive robot returns.

Spin "correction" and sim-to-sim adapter layers The simulation for paddle rubber has two sets of physical parameters, one for topspin and another for underspin. This causes a significant discrepancy between simulation and reality when using LLCs for topspin balls. We developed two solutions to mitigate this issue: topspin correction and sim-to-real adapter layers. First, we fine-tuned an LLC in simulation, dynamically selecting the appropriate ball-spin-dependent solver parameters, with additional rewards for low net clearance and target joint angle. This reduced the gap for specialized skills and increased ball return speed. However, a gap persisted for generalized skills on high topspin balls. To address this, we augmented the topspin-corrected policy with a FiLM layer Perez et al. (2018) with with 2.8k parameters and trained the adapter using just topspin balls for 5k steps. This closed the sim-to-real gap while preserving underspin return ability. Similar techniques could be applied to heavy underspin or side spin, but we leave this for future work.

Iteratively grounding the training task in the real world A seed dataset of 40 minutes of human vs. human play was collected along with 480 varied ball throws from a ball thrower. The sequence of ball positions was segmented into trajectories consisting of single ball hits and an offline optimization process (see Abeyruwan et al. (2023b)) was used to extract the initial ball state —– position, velocity, and angular velocity —– from each trajectory such that a simulated ball trajectory starting

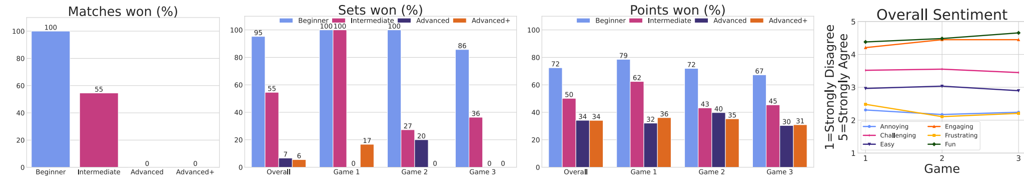

Figure 4: Match statistics and player sentiment measured by responses to "To what degree do these words describe your experience with playing table tennis with this robot?" on a five point Likert scale.

at that state matches the real ball trajectory as closely as possible. The output of this process resulted in a dataset of 2.6k initial ball states. An independent initial serving dataset of 0.9k balls was gathered separately. We extracted initial ball states from the serving trajectories using optimization methods described in Triggs et al. (2000).

Policies were trained in simulation with the objective of returning all balls in the dataset. During training, we sampled a ball state from the dataset, added small random perturbations, and validated the resulting trajectory. We then initialized the MuJoCo internal state with the ball state and started an episode. Since no training cycles were expended on unrealistic balls, model capacity was used more effectively, leading to faster training and higher return rates compared to the approach in Abeyruwan et al. (2023b). The resulting policies were deployed zero-shot to the real-world and evaluated against human opponents. Following the process outlined above, all evaluations were converted into a set of initial ball states and added to the dataset.

This iterative cycle of training models in simulation on the latest dataset, evaluating it in the real world, and using the annotated evaluation data to extend the dataset, can be repeated as many times as needed. We completed 7 cycles for rally balls and 2 cycles for serving balls over the course of 3 months with over 50 different human opponents, leading to a final dataset size of 14.2k initial ball states for rallies and 3.4k for serves. A summary of the dataset evolution is presented in Figure 3. One advantage of this approach is if the policy is repeatedly evaluated against diverse opponents, gaps in capabilities are automatically identified and filled. Performance had not plateaued after 7 cycles and we think further cycles could have continued to yield improvements.

Two further modifications to the training data distribution were important for boosting performance. (1) Reflecting the data along the y axis which doubled the final dataset size to 28k ball states. (2) Manually segmenting the dataset into 7 non-mutually exclusive categories — Fast, Normal speed, Slow, Topspin, No spin, Underspin, and Lob. During training, balls were selected by first sampling a category inversely proportional to the policy's return rate within that category, then an initial ball state was sampled uniformly from within that category. This focused training on harder categories while still maintaining performance on "easier" balls within those categories and across all categories.

3 Experiments and results

User study design To evaluate the skill level of our agent, we ran competitive matches against 29 unseen table tennis players of varying skill levels – beginner, intermediate, advanced, and advanced+ as determined by a survey and evaluation by a professional table tennis coach (who is also an author on this paper, henceforth referred to as "the coach"). The robot and human played three games, in which the first player to reach 11 points by a margin of two points (or a score of 20 points) won the game. The player that won the majority of games won the match. Unlike a real "best-of-three" match, all three games were played to ensure consistent data among participants. The coach acted as a referee to determine scoring and rules violations. Human players were given a minimum two-minute break between games to rest and fill out a short survey.

One major deviation from normal table tennis rules is that the robot cannot serve the ball. Serving conveys a strong advantage Katsikadelis et al. (2013) and is thus typically rotated every two points. To compensate for this limitation the human cannot win or lose points on the serve; the robot must return the ball and then points may be scored. This rule did lead to some more skilled players repeatedly attempting risky serves however we felt this was a necessary compromise to accommodate players of lower skill who may not be used to official serving rules. Two other limitations of the robot were accounted for. If the robot entered a protective stop state, the point was considered a "let" (no one scores). Similarly if the ball was hit very high (roughly 2 meters above the table) the point was also a let due to the limited field of view of the cameras. Applications of all rules were up to the referee's discretion.

Match results Figure 4 breaks down the matches between the humans and the robot. Overall the robot was solidly in the middle of the participants, winning 45% of matches, 46% of games, and 49% of points. When we break down matches by skill level, a clear pattern emerges. The robot won 100% of matches against beginner opponents, 55% of matches against intermediate opponents, and no matches against more skilled opponents. The implication is that the robot's skill level is intermediate; it can easily beat the previous skill level, is unable to win against higher skill levels and has roughly even odds to win against this skill level. That is not to say that the robot completely dominates or is dominated by other skills levels. Looking at the breakdown of points scored, the robot won 72% of points against beginners, 50% of points against intermediate players, and 34% of points against advanced and advanced+ players. Thus, the robot can still provide an interesting game to all levels of skills.

In addition to the quantitative match results, we also wanted to understand the qualitative side of this study; what was it actually like to play against a robot? Table tennis already has many so-called "robots" to aid in training, but these are essentially ball launchers whereas our system has the potential to be more dynamic, is better able to mimic the playstyles of a human, and carry on a full game. Analyzing the post-game surveys we see that most players did not employ a specific strategy in game 1 or were mostly focused on probing the robot's capabilities. During the second and third games, skilled players were able to identify gaps in the robot's capabilities, which correlated with higher win rates: players that mentioned "downspin", "backspin","chops", or "underspin" (a known weakness in the serving policies) in their game 2 and 3 comments were significantly more likely to have won their match ($p < 0.05$) and also to be of a higher skill level ($p < 0.001$).

We also wanted to ensure that playing with the robot was actually something people would want to do. Based on player feedback, we think this goal was achieved. Figure 4 (right) shows that across all skill groups players agreed that playing with the robot was "fun" and "engaging" based on a five point Likert scale. Novelty may play some role in this assessment, but the score tends to increase slightly over games and when players were offered additional time to freely play with the robot, 26/29 of them accepted and played for a mean of 4:06 and median of 5:00 out of a maximum of five minutes, implying that there is some lasting appeal to playing with the robot. Additionally, when asked "Would you be interested in playing with this robot again?", on a scale of one to five, the average response was 4.87 and the median response was 5.

HLC strategy analysis During each match, the HLC adapts to each opponent by learning numerical *preferences* (H-values) for the LLCs based on their online performance. The change in H-values during a match measures the extent of adaptation. For the forehand LLCs we consistently observe large changes in H-values of +/- 50% or more, and this trend holds across skill levels. However for the backhand LLCs the change in H-values was much smaller and often just a few percentage points. This indicates the HLC adapted when it played a forehand style but not the backhand. Qualitatively this is consistent with the observation from the coach that the backhand play was not at the level of the forehand during the matches. Since all matches began with the same initial H-values, the final H-values can be compared across skill groups to assess if the strategy differed. While there were some LLCs that were greatly preferred regardless of skill group, there were three LLCs that were preferred for beginners and one LLC that was preferred for the intermedate-advanced+ groups, indiciating that the HLC was able to adapt differently to different players.

4 Conclusions, Discussion and Future Work

We present the first robot agent capable of playing an interactive sport with humans at human level, representing a milestone in robot learning and control. It is a small step towards a long-standing goal in robotics of achieving human-level performance on many useful real world tasks.

Despite achieving amateur human-level performance, several limitations exist in our agent such as struggling against very low balls (due to safety constraints to avoid collision with the table), high balls (above the field of view of the cameras), and fast balls (due to system latency and lack of data). Additionally, it cannot detect the spin amount accurately and is limited to a short pip paddle that is easier to model. Advanced and advanced+ players were able to find and exploit these holes. We hypothesize our iterative learning method would fill the gaps and adapt to these players with more training rounds, within the physical capabilities of the robot. Further, the limitations may be addressed by exploring predictive models of ball trajectories, self-play techniques, learned reset pose or removing the reset, sophisticated collision avoidance algorithms, etc.

We also hope our research makes useful contributions beyond table tennis. Four aspects have broader implications:

(1) **Hierarchical policy** A library of low level skills (each specialized over a common model) and a high level controller that understands their strengths (via skill descriptors) and orchestrates them is a promising paradigm for efficient training on complex multi-task problems.

(2) **Sim-real synergy and iterative train-eval flywheel** We train in sim and deploy in real. Evaluation data is then added to the task distribution in sim. This enables automatic curriculum building and efficient continuous learning while bridging the sim-real gap from a task-distribution perspective. We believe this hybrid method is a fruitful area for future research.

(3) **Real time adaptation** Our agent tracks the human's strengths and also updates each of its own skills' performance online. This approach of online modeling of the agent and opponent's capabilities, and choosing the best suited skill for that context allows the agent to be robust and adapt efficiently to distribution shifts.

(4) **System design** To develop capable and robust controllers for complex real world tasks, system design may be as important as the algorithms, policy architectures and datasets. Every aspect of the system went through multiple rounds of optimization and redesign. This played a central role in the robustness and sim-to-real performance of the controller sustained over hours of gameplay.

These four components may help in building generalist robots that are capable of performing useful tasks at human-level, and interacting with humans in the real world.

REFERENCES

Globe 889 short pips-out table tennis rubber without sponge, 2024. https://shorturl.at/DQTTV [Last Accessed: (08/01/2024)].

Saminda Abeyruwan, Alex Bewley, Nicholas Matthew Boffi, Krzysztof Marcin Choromanski, David B D'Ambrosio, Deepali Jain, Pannag R Sanketi, Anish Shankar, Vikas Sindhwani, Sumeet Singh, et al. Agile catching with whole-body mpc and blackbox policy learning. In *Learning for Dynamics and Control Conference*, pp. 851–863. PMLR, 2023a.

Saminda Wishwajith Abeyruwan, Laura Graesser, David B D'Ambrosio, Avi Singh, Anish Shankar, Alex Bewley, Deepali Jain, Krzysztof Marcin Choromanski, and Pannag R Sanketi. i-sim2real: Reinforcement learning of robotic policies in tight human-robot interaction loops. In *Conference on Robot Learning*, pp. 212–224. PMLR, 2023b.

Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan, and Andy Zeng. Do as i can and not as i say: Grounding language in robotic affordances. In *arXiv preprint arXiv:2204.01691*, 2022.

Russell L Andersson. *A robot ping-pong player*, volume 988. MIT press Cambridge, Massachusetts, 1988.

R. C. Arkin. *Behavior-Based Robotics*. MIT Press, 1998.

Sven Behnke, Michael Schreiber, Jorg Stuckler, Reimund Renner, and Hauke Strasdat. See, walk, and kick: Humanoid robots start to play soccer. In *2006 6th IEEE-RAS International Conference on Humanoid Robots*, pp. 497–503. IEEE, 2006.

Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, sep 1975. ISSN 0001-0782. doi: 10.1145/361002.361007.

John Billingsley. Robot ping pong. *Practical Computing*, 1983.

Peter Blank, Benjamin H. Groh, and Björn M. Eskofier. Ball speed and spin estimation in table tennis using a racket-mounted inertial sensor. In Seungyon Claire Lee, Leila Takayama, Khai N. Truong, Jennifer Healey, and Thomas Ploetz (eds.), *ISWC*, pp. 2–9. ACM, 2017. ISBN 978-1-4503-5188-1.

R. Brooks. A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation*, 2(1):14–23, 1986. doi: 10.1109/JRA.1986.1087032.

D. Büchler, S. Guist, R. Calandra, V. Berenz, B. Schölkopf, and J. Peters. Learning to play table tennis from scratch using muscular robots. *IEEE Transactions on Robotics (T-RO)*, 38(6):3850–3860, 2022. doi: 10.1109/TRO.2022.3176207.

Ken Caluwaerts, Atil Iscen, J Chase Kew, Wenhao Yu, Tingnan Zhang, Daniel Freeman, Kuang-Huei Lee, Lisa Lee, Stefano Saliceti, Vincent Zhuang, et al. Barkour: Benchmarking animal-level agility with quadruped robots. *arXiv preprint arXiv:2305.14654*, 2023.

Micah Carroll, Rohin Shah, Mark K. Ho, Thomas L. Griffiths, Sanjit A. Seshia, Pieter Abbeel, and Anca D. Dragan. On the utility of learning about humans for human-ai coordination. *CoRR*, abs/1910.05789, 2019. URL `http://arxiv.org/abs/1910.05789`.

Letian Chen, Rohan Paleja, and Matthew Gombolay. Learning from suboptimal demonstration via self-supervised reward regression. In *Conference on robot learning*, pp. 1262–1277. PMLR, 2021.

Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. Extreme parkour with legged robots. *arXiv preprint arXiv:2309.14341*, 2023.

Open X-Embodiment Collaboration, Abby O'Neill, Abdul Rehman, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, Albert Tung, Alex Bewley, Alex Herzog, Alex Irpan, Alexander Khazatsky, Anant Rai, Anchit Gupta, Andrew Wang, Andrey Kolobov, Anikait Singh, Animesh Garg, Aniruddha Kembhavi, Annie Xie, Anthony Brohan, Antonin Raffin, Archit Sharma, Arefeh Yavary, Arhan Jain, Ashwin Balakrishna, Ayzaan Wahid, Ben Burgess-Limerick, Beomjoon Kim, Bernhard Schölkopf, Blake Wulfe, Brian Ichter, Cewu Lu, Charles Xu, Charlotte Le, Chelsea Finn, Chen Wang, Chenfeng Xu, Cheng Chi, Chenguang Huang, Christine Chan, Christopher Agia, Chuer Pan, Chuyuan Fu, Coline Devin, Danfei Xu, Daniel Morton, Danny Driess, Daphne Chen, Deepak Pathak, Dhruv Shah, Dieter Büchler, Dinesh Jayaraman, Dmitry Kalashnikov, Dorsa Sadigh, Edward Johns, Ethan Foster, Fangchen Liu, Federico Ceola, Fei Xia, Feiyu Zhao, Felipe Vieira Frujeri, Freek Stulp, Gaoyue Zhou, Gaurav S. Sukhatme, Gautam Salhotra, Ge Yan, Gilbert Feng, Giulio Schiavi, Glen Berseth, Gregory Kahn, Guangwen Yang, Guanzhi Wang, Hao Su, Hao-Shu Fang, Haochen Shi, Henghui Bao, Heni Ben Amor, Henrik I Christensen, Hiroki Furuta, Homer Walke, Hongjie Fang, Huy Ha, Igor Mordatch, Ilija Radosavovic, Isabel Leal, Jacky Liang, Jad Abou-Chakra, Jaehyung Kim, Jaimyn Drake, Jan Peters, Jan Schneider, Jasmine Hsu, Jeannette Bohg, Jeffrey Bingham, Jeffrey Wu, Jensen Gao, Jiaheng Hu, Jiajun Wu, Jialin Wu, Jiankai Sun, Jianlan Luo, Jiayuan Gu, Jie Tan, Jihoon Oh, Jimmy Wu, Jingpei Lu, Jingyun Yang, Jitendra Malik, João Silvério, Joey Hejna, Jonathan Booher, Jonathan Tompson, Jonathan Yang, Jordi Salvador, Joseph J. Lim, Junhyek Han, Kaiyuan Wang, Kanishka Rao, Karl Pertsch, Karol Hausman, Keegan Go, Keerthana Gopalakrishnan, Ken Goldberg, Kendra Byrne, Kenneth Oslund, Kento Kawaharazuka, Kevin Black, Kevin Lin, Kevin Zhang, Kiana Ehsani, Kiran Lekkala, Kirsty Ellis, Krishan Rana, Krishnan Srinivasan, Kuan Fang, Kunal Pratap Singh, Kuo-Hao Zeng, Kyle Hatch, Kyle Hsu, Laurent Itti, Lawrence Yunliang Chen, Lerrel Pinto, Li Fei-Fei, Liam Tan, Linxi "Jim" Fan, Lionel Ott, Lisa Lee, Luca Weihs, Magnum Chen, Marion Lepert, Marius Memmel, Masayoshi Tomizuka, Masha Itkina, Mateo Guaman Castro, Max Spero, Maximilian Du, Michael Ahn, Michael C. Yip, Mingtong Zhang, Mingyu Ding, Minho Heo, Mohan Kumar Srirama, Mohit Sharma, Moo Jin Kim, Naoaki Kanazawa, Nicklas Hansen, Nicolas Heess, Nikhil J Joshi, Niko Suenderhauf, Ning Liu, Norman Di Palo, Nur Muhammad Mahi Shafiullah, Oier Mees, Oliver Kroemer, Osbert Bastani, Pannag R Sanketi, Patrick "Tree" Miller, Patrick Yin, Paul Wohlhart, Peng Xu, Peter David Fagan, Peter Mitrano, Pierre Sermanet, Pieter Abbeel, Priya Sundaresan, Qiuyu Chen, Quan Vuong, Rafael Rafailov, Ran Tian, Ria Doshi, Roberto Mart'in-Mart'in, Rohan Baijal, Rosario Scalise, Rose Hendrix, Roy Lin, Runjia Qian, Ruohan Zhang, Russell Mendonca, Rutav Shah, Ryan Hoque, Ryan Julian, Samuel Bustamante, Sean Kirmani, Sergey Levine, Shan Lin, Sherry Moore, Shikhar Bahl, Shivin Dass, Shubham Sonawani, Shuran Song, Sichun Xu, Siddhant Haldar, Siddharth Karamcheti, Simeon Adebola, Simon Guist, Soroush Nasiriany, Stefan Schaal, Stefan Welker, Stephen Tian, Subramanian Ramamoorthy, Sudeep Dasari, Suneel Belkhale, Sungjae Park, Suraj Nair, Suvir Mirchandani, Takayuki Osa, Tanmay Gupta, Tatsuya Harada, Tatsuya Matsushima, Ted Xiao, Thomas Kollar, Tianhe Yu, Tianli Ding, Todor Davchev, Tony Z. Zhao, Travis Armstrong, Trevor Darrell, Trinity Chung, Vidhi Jain, Vincent Vanhoucke, Wei Zhan, Wenxuan Zhou, Wolfram Burgard, Xi Chen, Xiangyu Chen, Xiaolong Wang, Xinghao Zhu, Xinyang Geng, Xiyuan Liu, Xu Liangwei, Xuanlin Li, Yansong Pang, Yao Lu, Yecheng Jason Ma, Yejin Kim, Yevgen Chebotar, Yifan Zhou, Yifeng Zhu, Yilin Wu, Ying Xu, Yixuan Wang, Yonatan Bisk, Yongqiang Dou, Yoonyoung Cho, Youngwoon Lee, Yuchen Cui, Yue Cao, Yueh-Hua Wu, Yujin Tang, Yuke Zhu, Yunchu Zhang, Yunfan Jiang, Yunshuang Li, Yunzhu Li, Yusuke Iwasawa, Yutaka Matsuo, Zehan Ma, Zhuo Xu, Zichen Jeff Cui, Zichen Zhang, Zipeng Fu, and Zipeng Lin. Open X-Embodiment: Robotic learning datasets and RT-X models. `https://arxiv.org/abs/2310.08864`, 2023.

David B D'Ambrosio, Navdeep Jaitly, Vikas Sindhwani, Ken Oslund, Peng Xu, Nevena Lazic, Anish Shankar, Tianli Ding, Jonathan Abelian, Erwin Coumans, Gus Kouretas, Thinh Nguyen, Justin Boyd, Atil Iscen, Reza Mahjourian, Vincent Vanhoucke, Alex Bewley, Yuheng Kuang, Michael Ahn, Deepali Jain, Satoshi Kataoka, Omar E Cortes, Pierre Sermanet, Corey Lynch, Pannag R Sanketi, Krzysztof Choromanski, Wenbo Gao, Juhana Kangaspunta, Krista Reymann, Grace Vesom, Sherry Q Moore, Avi Singh, Saminda W Abeyruwan, and Laura Graesser. Robotic Table Tennis: A Case Study into a High Speed Learning System. In *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023. doi: 10.15607/RSS.2023.XIX.006.

Christian Daniel, Gerhard Neumann, Oliver Kroemer, and Jan Peters. Hierarchical relative entropy policy search. *Journal of Machine Learning Research*, 17(93):1–50, 2016.

Tianli Ding, Laura Graesser, Saminda Abeyruwan, David B D'Ambrosio, Anish Shankar, Pierre Sermanet, Pannag R Sanketi, and Corey Lynch. GoalsEye: Learning High Speed Precision Table Tennis on a Physical Robot. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 10780–10787. IEEE, 2022.

Zipeng Fu, Tony Z. Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. In *arXiv*, 2024.

Wenbo Gao, Laura Graesser, Krzysztof Choromanski, Xingyou Song, Nevena Lazic, Pannag Sanketi, Vikas Sindhwani, and Navdeep Jaitly. Robotic table tennis with model-free reinforcement learning. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5556–5563, 2020. doi: 10.1109/IROS45743.2020.9341191.

Yapeng Gao, Jonas Tebbe, Julian Krismer, and Andreas Zell. Markerless racket pose detection and stroke classification based on stereo vision for table tennis robots. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*, pp. 189–196, 2019. doi: 10.1109/IRC.2019.00036.

Thomas Gossard, Julian Krismer, Andreas Ziegler, Jonas Tebbe, and Andreas Zell. Table tennis ball spin estimation with an event camera, 2024. URL `https://arxiv.org/abs/2404.09870`.

Simon Guist, Jan Schneider, Hao Ma, Le Chen, Vincent Berenz, Julian Martus, Heiko Ott, Felix Grüninger, Michael Muehlebach, Jonathan Fiene, Bernhard Schölkopf, and Dieter Büchler. Safe accurate at speed with tendons: A robot arm for exploring dynamic motion, 2024. URL `https://arxiv.org/abs/2307.02654`.

Tuomas Haarnoja, Ben Moran, Guy Lever, Sandy H. Huang, Dhruva Tirumala, Jan Humplik, Markus Wulfmeier, Saran Tunyasuvunakool, Noah Y. Siegel, Roland Hafner, Michael Bloesch, Kristian Hartikainen, Arunkumar Byravan, Leonard Hasenclever, Yuval Tassa, Fereshteh Sadeghi, Nathan Batchelor, Federico Casarini, Stefano Saliceti, Charles Game, Neil Sreendra, Kushal Patel, Marlon Gwira, Andrea Huber, Nicole Hurley, Francesco Nori, Raia Hadsell, and Nicolas Heess. Learning agile soccer skills for a bipedal robot with deep reinforcement learning. *Science Robotics*, 9(89), 2024. doi: 10.1126/scirobotics.adi8022.

Hideaki Hashimoto, Fumio Ozaki, and Kuniji Osuka. Development of a pingpong robot system using 7 degrees of freedom direct drive arm. In *IECON'87: Industrial Applications of Robotics & Machine Vision*, volume 856, pp. 608–615. SPIE, 1987.

Hengyuan Hu, Adam Lerer, Alex Peysakhovich, and Jakob Foerster. "other-play" for zero-shot coordination. In *International Conference on Machine Learning*, pp. 4399–4410. PMLR, 2020.

Yanlong Huang, Bernhard Schölkopf, and Jan Peters. Learning optimal striking points for a ping-pong playing robot. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4587–4592. IEEE, 2015.

Yifeng Jiang, Tingnan Zhang, Daniel Ho, Yunfei Bai, C. Karen Liu, Sergey Levine, and Jie Tan. Simgan: Hybrid simulator identification for domain adaptation via adversarial reinforcement learning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2884–2890. IEEE Press, 2021. doi: 10.1109/ICRA48506.2021.9561731.

Michail Katsikadelis, Theophilos Pilianidis, and Nikolaos Mantzouranis. The interaction between serves and match winning in table tennis players in the london 2012 olympic games. In *Book of abstracts of the 8th international table tennis federation sports science congress–the 3rd world congress of science and racket sports*, pp. 77–79, 2013.

Elia Kaufmann, Leonard Bauersfeld, Antonio Loquercio, Matthias Mueller, Vladlen Koltun, and Davide Scaramuzza. Champion-level drone racing using deep reinforcement learning. *Nature*, 620:982–987, 08 2023. doi: 10.1038/s41586-023-06419-4.

Mitsuo Kawato, Kazunori Furukawa, and Ryoji Suzuki. A hierarchical neural-network model for control and learning of voluntary movement. *Biological cybernetics*, 57:169–185, 1987.

Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda, and Eiichi Osawa. Robocup: The robot world cup initiative. In *Proceedings of the First International Conference on Autonomous Agents*, AGENTS '97, pp. 340–347, New York, NY, USA, 1997. Association for Computing Machinery. ISBN 0897918770. doi: 10.1145/267658.267738.

John Knight and David Lowery. Pingpong-playing robot controlled by a microcomputer. *Microprocessors and Microsystems - Embedded Hardware Design*, 1986.

Okan Koç, Guilherme Maeda, and Jan Peters. Online optimal trajectory generation for robot table tennis. *Robotics and Autonomous Systems*, 105:121–137, 2018.

Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. Rma: Rapid motor adaptation for legged robots. 2021.

Asai Kyohei, Nakayama Masamune, and Yase Satoshi. The ping pong robot to return a ball precisely trajectory prediction and racket control for spinning balls. 2019. URL `https://api.semanticscholar.org/CorpusID:214698536`.

Kuan-Hui Lee, German Ros, Jie Li, and Adrien Gaidon. Spigan: Privileged adversarial learning from simulation. In *International Conference on Learning Representations*, 2018.

Chenhao Li, Marin Vlastelica, Sebastian Blaes, Jonas Frey, Felix Grimminger, and Georg Martius. Learning agile skills via adversarial imitation of rough partial demonstrations. In Karen Liu, Dana Kulic, and Jeff Ichnowski (eds.), *Proceedings of The 6th Conference on Robot Learning*, volume 205 of *Proceedings of Machine Learning Research*, pp. 342–352. PMLR, 14–18 Dec 2023. URL `https://proceedings.mlr.press/v205/li23b.html`.

Chunfang Liu, Yoshikazu Hayakawa, and Akira Nakashima. Racket control for a table tennis robot to return a ball. *SICE Journal of Control, Measurement, and System Integration*, 6:259–266, 07 2013. doi: 10.9746/jcmsi.6.259.

M. Matsushima, T. Hashimoto, and F. Miyazaki. Learning to the robot table tennis task-ball control & rally with a human. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, volume 3, pp. 2962–2969 vol.3, 2003. doi: 10.1109/ICSMC.2003.1244342.

Shotaro Mori, Kazutoshi Tanaka, Satoshi Nishikawa, Ryuma Niiyama, and Yasuo Kuniyoshi. High-speed humanoid robot arm for badminton using pneumatic-electric hybrid actuators. *IEEE Robotics and Automation Letters*, 4(4):3601–3608, 2019. doi: 10.1109/LRA.2019.2928778.

Katharina Muelling, Jens Kober, and Jan Peters. Learning table tennis with a Mixture of Motor Primitives. *IEEE-RAS Humanoids*, 2010.

Katharina Muelling, Abdeslam Boularias, Betty Mohler, Bernhard Schölkopf, and Jan Peters. Learning strategies in table tennis using inverse reinforcement learning. *Biol. Cybern.*, 108(5):603–619, oct 2014. ISSN 0340-1200. doi: 10.1007/s00422-014-0599-1.

Katharina Mülling, Jens Kober, Oliver Kroemer, and Jan Peters. Learning to select and generalize striking movements in robot table tennis. *The International Journal of Robotics Research*, 32(3): 263–279, 2013.

Akira Nakashima, Yuki Ogawa, Chunfang Liu, and Yoshikazu Hayakawa. Robotic table tennis based on physical models of aerodynamics and rebounds. *2011 IEEE International Conference on Robotics and Biomimetics*, 2011.

Aäron van den Oord, Nal Kalchbrenner, Oriol Vinyals, Lasse Espeholt, Alex Graves, and Koray Kavukcuoglu. Conditional image generation with pixelcnn decoders. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pp. 4797–4805, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J Andrew Bagnell, Pieter Abbeel, Jan Peters, et al. An algorithmic perspective on imitation learning. *Foundations and Trends® in Robotics*, 7(1-2): 1–179, 2018.

Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 3803–3810. IEEE, 2018.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron C. Courville. Film: Visual reasoning with a general conditioning layer. In Sheila A. McIlraith and Kilian Q. Weinberger (eds.), *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 3942–3951. AAAI Press, 2018. doi: 10.1609/AAAI.V32I1.11671.

Tadej Petric, Luka Peternel, Andrej Gams, Bojan Nemec, and Leon Zlajpah. Navigation methods for the skiing robot. *International Journal of Humanoid Robotics*, 10, 01 2012. doi: 10.1142/S0219843613500291.

Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

Thomas Röfer, Tim Laue, Arne Hasselbring, Jo Lienhoop, Yannik Meinken, and Philip Reichenberg. B-human 2022 – more team play with less communication. In Amy Eguchi, Nuno Lau, Maike Paetzel-Prüsmann, and Thanapat Wanichanon (eds.), *RoboCup 2022: Robot World Cup XXV*, pp. 287–299, Cham, 2023. Springer International Publishing. ISBN 978-3-031-28469-4.

Julio K. Rosenblatt and Charles E. Thorpe. *A Behavior-based Architecture for Mobile Navigation*, pp. 19–32. Springer US, Boston, MA, 1997. ISBN 978-1-4615-6325-9. doi: 10.1007/978-1-4615-6325-9_2.

Gerhard Schweitzer and Jianyong Wen. Where neural nets make sense in robotics. In *Prerational Intelligence: Adaptive Behavior and Intelligent Systems Without Symbols and Logic, Volume 1, Volume 2 Prerational Intelligence: Interdisciplinary Perspectives on the Behavior of Natural and Artificial Systems, Volume 3*, pp. 530–560. Springer, 1994.

Nitish Sontakke, Hosik Chae, Sangjoon Lee, Tianle Huang, Dennis W. Hong, and Sehoon Hal. Residual physics learning and system identification for sim-to-real transfer of policies on buoyancy assisted legged robots. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 392–399, 2023. doi: 10.1109/IROS55552.2023.10342062.

Simon Stepputtis, Joseph Campbell, Mariano Phielipp, Stefan Lee, Chitta Baral, and Heni Ben Amor. Language-conditioned imitation learning for robot manipulation tasks. *Advances in Neural Information Processing Systems*, 33:13139–13150, 2020.

Peter Stone, Richard S Sutton, and Gregory Kuhlmann. Reinforcement learning for robocup soccer keepaway. *Adaptive Behavior*, 13(3):165–188, 2005.

Peter Stone, Gal Kaminka, Sarit Kraus, and Jeffrey Rosenschein. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 24, pp. 1504–1509, 2010.

DJ Strouse, Kevin McKee, Matt Botvinick, Edward Hughes, and Richard Everett. Collaborating with humans without human data. *Advances in Neural Information Processing Systems*, 34: 14502–14515, 2021.

Yichao Sun, Rong Xiong, Qiuguo Zhu, Jun Wu, and Jian Chu. Balance motion generation for a humanoid robot playing table tennis. In *2011 11th IEEE-RAS International Conference on Humanoid Robots*, pp. 19–25, 2011. doi: 10.1109/Humanoids.2011.6100826.

Vincenzo Suriani, Emanuele Musumeci, Daniele Nardi, and Domenico Daniele Bloisi. Play everywhere: A temporal logic based game environment independent approach for playing soccer with robots. In Cédric Buche, Alessandra Rossi, Marco Simões, and Ubbo Visser (eds.), *RoboCup 2023: Robot World Cup XXVI*, pp. 3–14, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-55015-7.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.

Jonas Tebbe, Yapeng Gao, Marc Sastre-Rienietz, and Andreas Zell. A Table Tennis Robot System Using an Industrial KUKA Robot Arm. *GCPR*, 2018.

Jonas Tebbe, Lukas Krauch, Yapeng Gao, and Andreas Zell. Sample-efficient reinforcement learning in robotic table tennis. In *2021 IEEE international conference on robotics and automation (ICRA)*, pp. 4171–4178. IEEE, 2021.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033, 2012. doi: 10.1109/IROS.2012.6386109.

Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, ICCV '99, pp. 298–372, London, UK, UK, 2000. Springer-Verlag. ISBN 3-540-67973-1.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

Sinuo Wang, Maëlic Neau, and Cédric Buche. Robonlu: Advancing command understanding with a novel lightweight bert-based approach for service robotics. In Cédric Buche, Alessandra Rossi, Marco Simões, and Ubbo Visser (eds.), *RoboCup 2023: Robot World Cup XXVI*, pp. 29–41, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-55015-7.

Zhikun Wang, Katharina Mülling, Marc Peter Deisenroth, Heni Ben Amor, David Vogt, Bernhard Schölkopf, and Jan Peters. Probabilistic movement modeling for intention inference in human–robot interaction. *The International Journal of Robotics Research*, 32(7):841–858, 2013.

Zhikun Wang, Abdeslam Boularias, Katharina Muelling, Bernhard Schölkopf, and Jan Peters. Anticipatory action selection for human-robot table tennis. *Artif. Intell.*, 2017.

Jimmy Wu, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon Rusinkiewicz, and Thomas Funkhouser. Tidybot: personalized robot assistance with large language models. *Auton. Robots*, 47(8):1087–1102, nov 2023. ISSN 0929-5593. doi: 10.1007/s10514-023-10139-z. URL https://doi.org/10.1007/s10514-023-10139-z.

Boling Yang, Xiangyu Xie, Golnaz Habibi, and Joshua R Smith. Competitive physical human-robot game play. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 242–246, 2021.

Boling Yang, Golnaz Habibi, Patrick Lancaster, Byron Boots, and Joshua Smith. Motivating physical activity via competitive human-robot interaction. In *Conference on Robot Learning*, pp. 839–849. PMLR, 2022.

Zulfiqar Zaidi, Daniel Martin, Nathaniel Belles, Viacheslav Zakharov, Arjun Krishna, Kin Man Lee, Peter Wagstaff, Sumedh Naik, Matthew Sklar, Sugju Choi, Yoshiki Kakehi, Ruturaj Patil, Divya Mallemadugula, Florian Pesce, Peter Wilson, Wendell Hom, Matan Diamond, Bryan Zhao, Nina Moorman, Rohan Paleja, Letian Chen, Esmaeil Seraj, and Matthew Gombolay. Athletic mobile manipulator system for robotic wheelchair tennis. *IEEE Robotics and Automation Letters*, 8 (4):2245–2252, April 2023. ISSN 2377-3774. doi: 10.1109/lra.2023.3249401. URL http://dx.doi.org/10.1109/LRA.2023.3249401.

Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 737–744, 2020. doi: 10.1109/SSCI47803.2020.9308468.

Yifeng Zhu, Yongsheng Zhao, Lisen Jin, Jun Wu, and Rong Xiong. Towards high level skill learning: Learn to return table tennis ball using monte-carlo based policy gradient method. In *2018 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pp. 34–41, 2018. doi: 10.1109/RCAR.2018.8621776.