



<http://www.diva-portal.org>

## Postprint

This is the accepted version of a paper presented at *Context-Awareness in Human-Robot Interaction: Approaches and Challenges, workshop at 2022 ACM/IEEE International Conference on Human-Robot Interaction*.

Citation for the original published paper:

Deichler, A., Wang, S., Alexanderson, S., Beskow, J. (2022)  
Towards Context-Aware Human-like Pointing Gestures with RL Motion Imitation  
In: (pp. 2022-).

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-313480>

# Towards Context-Aware Human-like Pointing Gestures with RL Motion Imitation

Anna Deichler\*

*Division of Speech, Music and Hearing  
KTH Royal Institute of Technology  
Stockholm, Sweden  
deichler@kth.se*

Siyang Wang\*

*Division of Speech, Music and Hearing  
KTH Royal Institute of Technology  
Stockholm, Sweden  
siyangw@kth.se*

Simon Alexanderson

*Division of Speech, Music and Hearing  
KTH Royal Institute of Technology  
Stockholm, Sweden  
simonal@kth.se*

Jonas Beskow

*Division of Speech, Music and Hearing  
KTH Royal Institute of Technology  
Stockholm, Sweden  
beskow@kth.se*

**Abstract**—Pointing is an important mode of interaction with robots. While large amounts of prior studies focus on recognition of human pointing, there is a lack of investigation into generating context-aware human-like pointing gestures, a shortcoming we hope to address. We first collect a rich dataset of human pointing gestures and corresponding pointing target locations with accurate motion capture. Analysis of the dataset shows that it contains various pointing styles, handedness, and well-distributed target positions in surrounding 3D space in both single-target pointing scenario and two-target point-and-place. We then train reinforcement learning (RL) control policies in physically realistic simulation to imitate the pointing motion in the dataset while maximizing pointing precision reward. We show that our RL motion imitation setup allows models to learn human-like pointing dynamics while maximizing task reward (pointing precision). This is promising for incorporating additional context in the form of task reward to enable flexible context-aware pointing behaviors in physically realistic environment while retaining human-likeness in pointing motion dynamics.

**Index Terms**—motion generation, pointing gesture, reinforcement learning, human-robot interaction, motion capture

## I. INTRODUCTION

There is an increasing number of studies on human-robot interaction using pointing gestures [1], [9], [21]. The emphasis has so far been put on recognizing pointing from human user and distinguishing the pointed object, while generation of pointing behavior in robots is less studied. Pointing as a functional gesture has two important aspects; legibility, or how accurate is the pointing [12], and motion dynamics, or the human-likeness of the pointing motion itself. The latter aspect is especially lacking in prior studies. Low human-likeness in pointing motion can hinder human perception of the robot intention. Moreover, humans point with varying dynamics to convey extra information besides just what is being pointed at.

\* equal contributions. This work was financed by Digital Futures, project “Advanced Adaptive Intelligent Systems” and Swedish Research Council, grant nr 2018-05409.

For example, a quick point and stop motion when referring to an object shows certainty, while a slower pointing with sub-movements (e.g. hovering around pointed object) indicates less confidence or certainty. The dynamics of pointing motion can also be interpreted as style factors that may convey information about individual personality.

In recent years, methods for generating naturalistic and believable gesture for virtual characters and social robots have advanced tremendously, thanks to new methods in probabilistic motion modelling using deep architectures and the availability of extensive motion capture datasets [3], [23]. However, such methods require a large amount of data (motion capture) and are not designed to adapt to unseen environments.

In this paper we propose to use a combination of data-driven and physically based reinforcement learning (RL) for communicative gestures. The reinforcement learning component allows the model to flexibly adapt itself to task rewards (e.g. pointing accuracy) during simulation training or real-world adaptation. We take inspiration from prior work in the domain of physically based character animation [18], [19], and discuss how these approaches may be applied to gestures that are communicative in nature and also dependent on the context of the environment. Our initial experiments show promise of RL-based motion imitation approach for generating dynamic, accurate, and human-like pointing gestures.

We also collect a new motion-captured pointing gesture dataset. We present summary statistics and visualization of target positions, handedness (left or right), and range of movement in the dataset. They show that the dataset contains well-distributed target positions and diverse range of movement, thus it has the potential to be used in several different tasks, such as generating pointing behaviors (this paper), or training a model to identify pointing gesture and pointed target position. We show that the arm of the actor is not precisely aligned with target position. This suggests that a simple metric like the alignment angle is *not* enough to, (1) be the only RL reward

in generating human-like pointing as our data analysis show that humans do not point "accurately" in terms of minimizing alignment angle, (2) be the only criteria in discriminating the object a user is pointing to.

## II. RELATED WORK

Pointing has been extensively studied in human-robot interaction. [20] has developed a human-robot gesture language for two-way communication in pick-and-place tasks, which included pointing. Prior studies have concentrated on the recognition of human pointing gestures [17], [8]. There are fewer studies on generation of pointing behavior in robots. [12] proposes a mathematical optimization approach to make legible pointing, i.e. disambiguation of closely situated objects. A more recent work [24] proposes a reinforcement learning approach with actor-critic setting to achieve legible pointing. An imperial study on making humanoid robot exhibit guide [13] employs a pointing generation sub-system. These studies mostly focus on pointing accuracy, and do not specifically address human-likeness of pointing behaviors from a motion dynamics perspective.

In human communication, deictic gestures (pointing) are ubiquitous, and one of the first means of communication developed in infancy [14]. Pointing helps us communicate more effectively and robustly, guide visual attention [6], and understand other people's intentions [5]. Pointing typically convey complementary information to the accompanying speech rather than redundant [7]. Studies on human pointing have also revealed that humans are able to point accurately with variable speed [2] and to point with variable trajectories [4]. This presents challenge for learning physically realistic and human-like pointing gestures in robots.

Gesture synthesis for virtual agents and social robots has recently shifted from rule-based systems employing instructions given in a markup language [16], [22] to data-driven methods generating gestures directly from speech acoustics or text [3], [23]. This development has increased naturalness and ease-of-use to the trade-off of losing ability to generate complementary gestures. This paper aims to bridge the gap by proposing a learning environment and framework to learn complementary aspects of speech and pointing gestures.

Reinforcement learning (RL) has been extensively used in robotics studies, since compared to other machine learning methods it allows the agent to learn from interacting with its environment, therefore it allows adaptability, compared to passive supervised machine learning settings [15]. In recent years, accurate physics engines enabled training RL agents in ever more realistic scenarios. However, it is challenging to produce natural motor behavior for physically simulated humanoids. Therefore recent works have explored a combination of motion capture based imitation learning with RL to achieve generalizable and human-like skill learning [18], [19], [10].

## III. MOTION CAPTURE DATA

A pointing gesture dataset was recorded in a optical motion capture studio equipped with 16 Optitrack Prime 41 cameras.

The actor wore a suit with 50 passive markers, a pair of Manus data gloves (hand motion capture) and a head-mounted iPhone 12 (face motion and voice capture). The motion capture system recorded the actor's joint positions, as well as the target positions in 3D space. Three different pointing tasks were executed: single target pointing, two targets selection and two targets moving (point-and-place). In each setting the targets were moved around in order to get a coverage of the surrounding space, while the actor was in a stationary position. Between each pointing task the actor returned to a neutral stance. The beginning and the end of the movement is defined by the pointing hand leaving and returning to the initial, downward position.

The dataset also includes recorded speech as the actor is pointing. Example utterances are, "I want that one" (in single-target pointing), and "Put this one there" (in two-target pointing and placing). The speech data is synchronized with mocap. This provides extra context for training models to generate pointing behaviour, and recognition of pointing timing, perhaps using speech as extra input to better identify the instant of pointing instead of relying on the motion itself.

### A. Data analysis

We focus our data analysis on the single target pointing task, since models in this study are trained on single target clips. The single target dataset is parsed into single arm pointing movement based on peaks in displacement on the sagittal plane. The resulting target position counts can be read in Table I. The total sum of single target positions in the dataset is 83. There are 32 left-handed and 51 right-handed pointing movements. The actor's dominant hand is right hand.

We analyse the pointing movements in terms of duration of pointing movement, duration of rise time, peak velocity, average velocity and endpoint precision. The velocity data is obtained from numerical differentiation of the  $x, y, z$  positional data from the markers placed on the left and right hands. The average velocity is calculated for the rise phase of the pointing motion. We present the result of our analysis on single target pointing movements in Table II. Overall, the table shows that there is clear difference between the dominant and non-dominant hands. The peak and mean velocities of the dominant arm are higher than the non-dominant arm and rise time is also lower. The overall duration is longer for the dominant hand, since the pointing was held out longer. On the top plot in Figure 1 we can see the averaged velocity profiles over the whole pointing movement for both hands. On the bottom plot the averaged pointing precision are depicted for both hands. We measure the pointing precision as a modified pointing angle term proposed in section IV-B. The maximum measure is 1 and can only be obtained when the arm is fully aligned with the pointed target (angle equals 0), however as seen in the plot, neither dominant or non-dominant hands achieve this measure or come close to achieving it. This shows that human pointing is not accurate in terms of alignment angle an extensively used metric in prior pointing gesture recognition and generation studies.

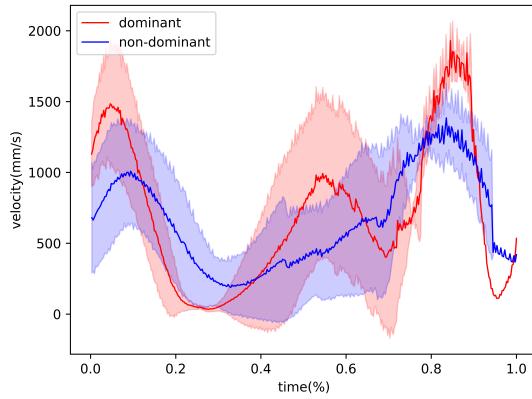


Fig. 1. Average hand velocity profiles and average pointing precision profiles from the single pointing task for dominant and non-dominant hands.

TABLE I  
SINGLE TARGET POSITION COUNTS

	Front		Back	
	Left	Right	Left	Right
Top	11	15	8	8
Bottom	14	12	7	8

We visualize<sup>1</sup> the target distribution in space in Figure 2 (single-target pointing). Different colors are given to quadrants divided by left/right, top/down, front/back. Positions are distributed evenly in each quadrant.

<sup>1</sup>All rendering is done using <https://github.com/xbpeng/DeepMimic>.

TABLE II  
SINGLE TARGET POINTING MOVEMENTS

	Both	Dominant	Non-dominant
Duration (s)	$2.4 \pm 1.05$	$2.61 \pm 1.13$	$2.11 \pm 0.86$
Rise time (s)	$0.71 \pm 0.31$	$0.70 \pm 0.3$	$0.73 \pm 0.32$
Peak velocity (mm/s)	$1706 \pm 392$	$1825 \pm 384$	$1541 \pm 340$
Mean velocity (mm/s)	$849 \pm 302$	$925 \pm 317$	$745 \pm 244$

<sup>a</sup>Unless noted otherwise, all statistics in this paper are reported as [mean] $\pm$ [sd].

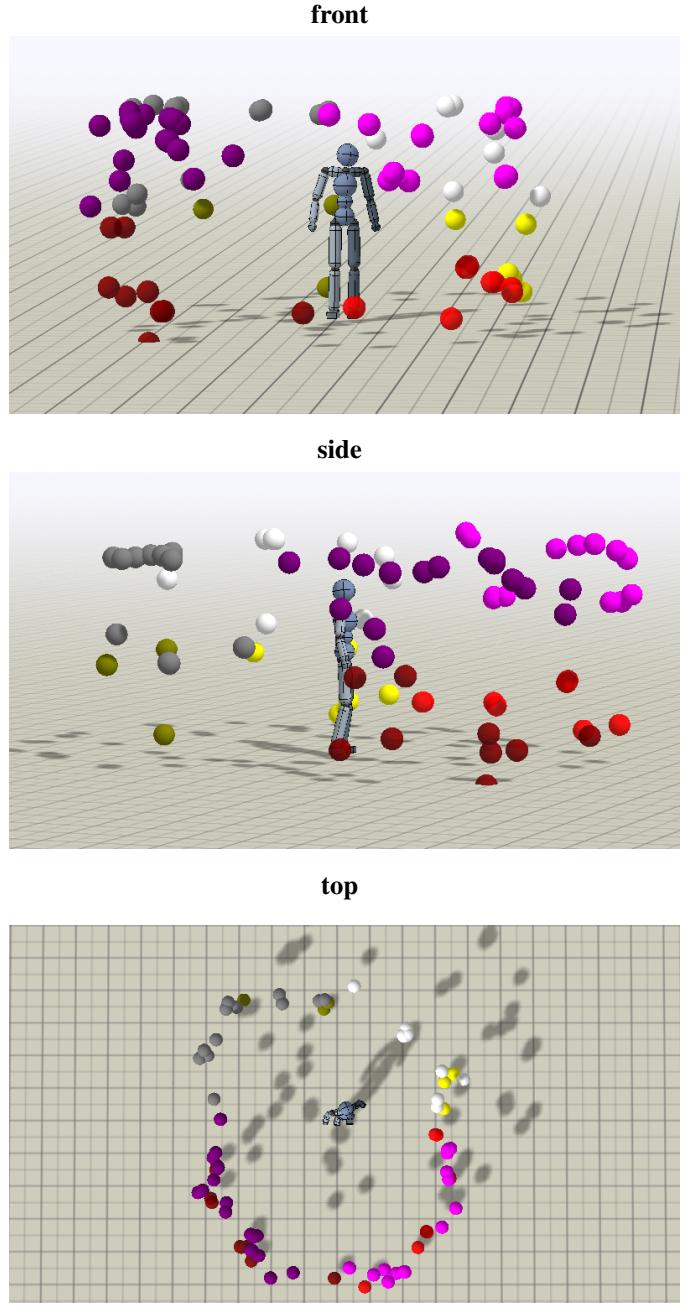


Fig. 2. Target position distribution in single-target pointing.

Target distribution for two-target pointing task is shown in Figure 3. The arrows connect two targets that are pointed in a two-target pointing sequence. Our dataset not only covers a wide range of target positions but also a wide range of two-target movement.

#### IV. EXPERIMENT

Our goal is to achieve human-like pointing in a physically simulated humanoid character that can point to any given target position in its surrounding 3D space. To this end, we use a reinforcement learning (RL) setup, where we aim to learn a control policy which can reproduce the dynamics of

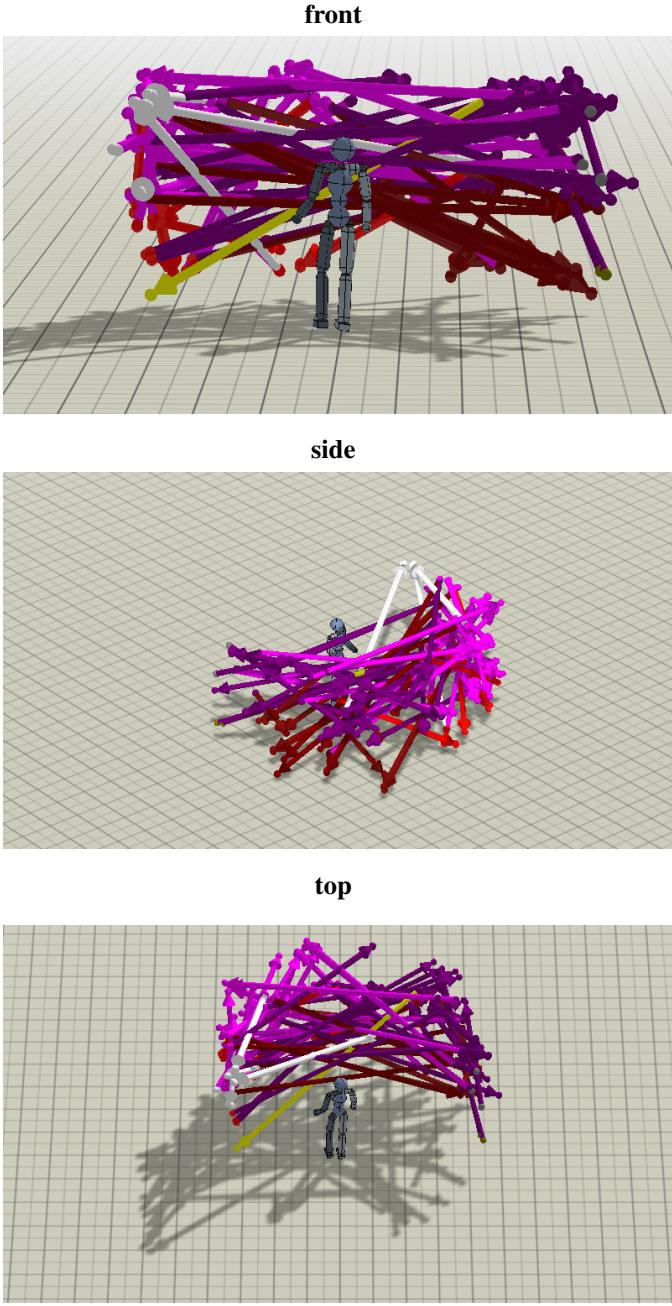


Fig. 3. Target position distribution in two-target pointing and placing. The arrow direction indicating the from- and to- targets. The colors indicating quadrants (same as Figure 2) of the end target positions.

the pointing movements from the motion capture dataset and complete the pointing task defined by the pointing reward function. At the same time, the policy should also be robust, namely without falling or other fault behaviors. The RL setup also allows future adaption of arbitrary additional context or modification of rewards to achieve desired style or level of pointing precision.

### A. Methodology

The learning model is based on Deepmimic (DM) [18] and Adversarial Motion Priors (AMP) [19]<sup>2</sup>. The model uses motion capture data based imitation and reinforcement learning to achieve human-like motion and to complete a specified task at the same time.

At each time step  $t$  the reward function is the combination of rewards from motion capture imitation  $r_t^I$  and the task reward  $r_t^G$  (discussed in section IV-B), weighted by pre-defined weights  $\omega^I, \omega^G$ .

$$r_t = \omega^I r_t^I + \omega^G r_t^G \quad (1)$$

For DM models, we keep the original composition of the imitation reward, which is a weighted combination of pose, velocity, end-effector and center-of-mass rewards

$$\omega^I r_t^I = \omega^p r_t^p + \omega^v r_t^v + \omega^e r_t^e + \omega^c r_t^c. \quad (2)$$

AMP replaces imitation rewards with a single neural-networks-based discriminator. We use same weights for reward terms as original papers [18], [19].

We test 4 configurations of DM models. First 3 are generic DM models with pointing target as extra input to the policy network, but they do not receive task reward. These three differ in the number of motion clips used in training, they are denoted as **DM-4**, **DM-16**, **DM-32** where the numbers denote the number of clips. We test an additional DM model **DM-16-wr** for which the task reward is given along with imitation reward. We test a single AMP model **AMP-fr**, here *fr* refers to "front raise" because we only give this model front pointing motion clips and all motion clips only contain the arm raising stage of the pointing but not retraction. We use this setup because AMP struggles with distinguishing between the raising and retraction phases of pointing gesture, a problem we hope to solve in future studies. We also restrict target positions for all models to front-facing targets only for easier training and more straightforward comparisons, but the methods can be directly applied to targets in the back.

### B. Reward functions for pointing

In order to achieve the objective of pointing, an appropriate reward function should be devised. In this paper, we design a geometric reward function that rewards the agent if the vector defined by its forearm is aligned with the vector connecting its end-effector and the target as seen on Figure 4.

$$\hat{\theta} = 1 - \frac{\angle(\vec{V}_{HT}, \vec{V}_{EH})}{\pi} \quad , \quad r_t^{Pt} = \frac{e^{\hat{\theta}} - 1}{e} \quad (3)$$

### C. Results

We compare the models on a separate set of 12 target positions outside of the training set (AMP training has seen these target positions but not the motions directly, so the comparison remains valid). To quantify overall pointing accuracy, we calculate reward statistics over the 12 pointing sequences

<sup>2</sup>Modified implementation from <https://github.com/xbpeng/DeepMimic>.

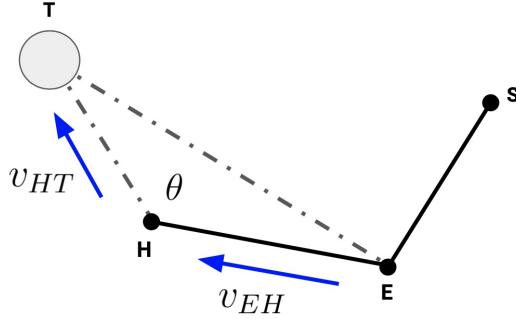


Fig. 4. Pointing reward is defined based on angle between vector from elbow joint (E) to hand joint (H) and vector from hand joint to target position (T) in 3D space.

for each model as shown in Table III. **AMP-fr** obtains highest  $r_{\max}$ , showing that it has learned to point more accurate than other models. **DM-32** has the highest pointing accuracy out of all DM models, which could be due to its higher number of training data.

We plot the angle rewards obtained by the models (along with the ground-truth motion) when starting at the same idle pose and given the same target position in Figure 5. Two sample clips are shown. The horizon of frames ends at the highest ground-truth motion reward. **DM-32** in Clip 1 is an example of a good model pointing behavior, where both the raising and retraction motions are smooth as seen by the smoothness of the reward curves. The pointing is also accurate as the highest reward at the peak of bell curve is actually higher than ground-truth motion. The plots also show that several models make undesirable jerky, vibration-like motion towards the pointing target.

TABLE III  
POINTING REWARDS OBTAINED BY MODELS

	$r_{\max}$	$r_{\min}$	$r_{\text{mean}}$
<b>DM-4</b>	$0.53 \pm 0.08$	$0.34 \pm 0.09$	$0.46 \pm 0.07$
<b>DM-16</b>	$0.52 \pm 0.08$	$0.37 \pm 0.06$	$0.46 \pm 0.07$
<b>DM-16-wr</b>	$0.44 \pm 0.08$	$0.34 \pm 0.07$	$0.39 \pm 0.07$
<b>DM-32</b>	$0.55 \pm 0.09$	$0.34 \pm 0.07$	$0.48 \pm 0.08$
<b>AMP-fr</b>	$0.58 \pm 0.05$	$0.35 \pm 0.05$	$0.47 \pm 0.04$

<sup>a</sup>Each metric is first calculated per test motion clip (12 in total) and then calculated summary statistics which are shown here.

TABLE IV  
MOTION SMOOTHNESS COMPARISON

	$\text{vel}_r (\text{r/t})$	$\text{acc}_r (\text{r/t})$	$\text{jerk}_r (\text{r/t})$
<b>DM-4</b>	$0.33 \pm 0.26$	$9.61 \pm 10.69$	$739.11 \pm 940.16$
<b>DM-16</b>	$1.11 \pm 0.28$	$84.74 \pm 19.06$	$9328.20 \pm 2349.19$
<b>DM-16-wr</b>	$0.94 \pm 0.16$	$81.19 \pm 17.74$	$8615.28 \pm 2449.04$
<b>DM-32</b>	$0.35 \pm 0.25$	$8.87 \pm 8.85$	$592.48 \pm 549.79$
<b>AMP-fr</b>	$0.47 \pm 0.20$	$7.72 \pm 3.06$	$350.85 \pm 130.22$

<sup>a</sup>Lower measures in all three metrics indicate smoother motion.

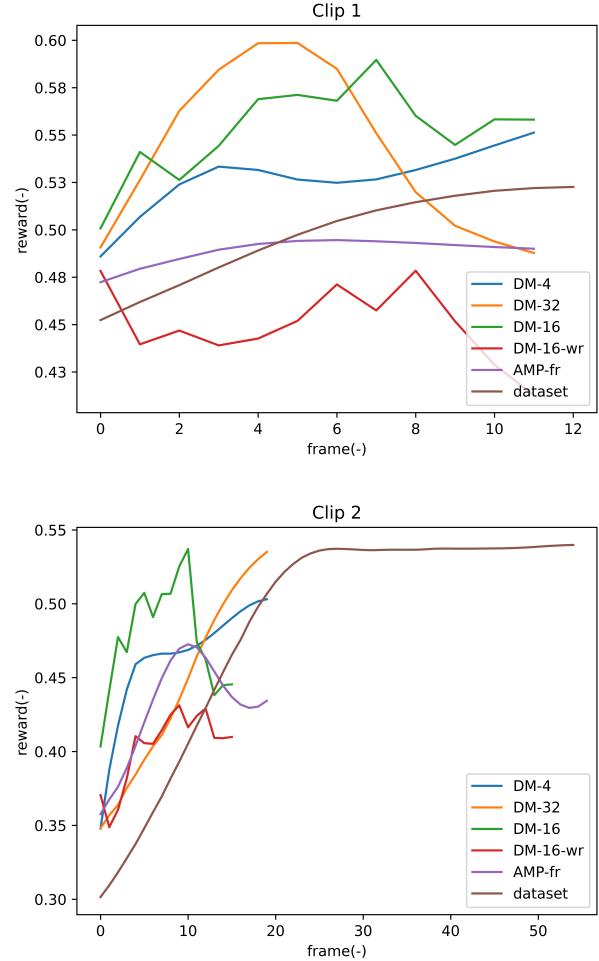


Fig. 5. Reward comparison for different settings

To further quantify difference in motion smoothness, we calculate the rate of change of angle reward  $r$  ( $\text{vel}_r$ ), acceleration ( $\text{acc}_r$ ), and jerk ( $\text{jerk}_r$ ). The results are shown in Table IV. **AMP-fr** obtains lowest measures in all three categories indicating that its generated motions are smoother than other models. This is consistent with our visual inspection. **DM-16** obtains highest measures in all three categories, which is also consistent with our observation that it is indeed unstable and constantly makes jerky motions in pointing arm towards the target. These results suggest that AMP is advantageous in obtaining smooth motion and should be tested more extensively in future studies.

## V. DISCUSSION

This paper presents initial results of learning pointing skills in a physically simulated agent using reinforcement learning. In future studies, our main focuses are incorporating additional context in the form of task reward and re-evaluating how the pointing task is defined, which includes modifying the data collection procedure, and testing different inputs to the learning system and designing more appropriate reward functions.

We also want to incorporate motion style using a custom style reward function.

The current setup aimed at reproducing the dynamics of pointing movements, which were recorded in a rigid setting. We plan to design an experimental setup, which reflects everyday communication scenarios more closely (e.g. target selection in games). On the algorithmic side, we also plan to extend the system to accept observational data (ego-centric images) and speech as input. Pointing gestures are most often accompanied by verbal communication, which can provide context and can help resolving disambiguities in the interlocutor's side in understanding the pointing gesture. Ego-centric images would allow us to learn hand-eye-coordination in the pointing task.

We also plan to re-evaluate the reward function design, especially how we can achieve context-aware pointing by incorporating additional context in the form of reward functions. Achieving the intended objective in reinforcement learning is largely determined by devising an appropriate reward function, however this is often a challenging task. In this paper we took a reward function based on simple geometrical relations between the target and the agent's end-effector. However, there are no clear geometric rules for guiding the production and interpretation of pointing gestures even in human-human communication and people often fail to determine the target location from pointing gestures [11], a finding also confirmed in analysis on our motion-captured pointing gesture dataset as the alignment angle between actor's arm and pointed target is not near 0. But humans can often refer to the object accurately with pointing in the interaction context. To this end we believe that, since pointing is a communicative signal, its success should also be measured by the recognition of the intended target. Thus adding context-aware recognition of pointed object to the learning loop could be key to a successful system. In future studies we also intend to make use of user studies to evaluate the generated pointing behavior in context. These evaluations could include disambiguation from closely situated targets.

## REFERENCES

- [1] Shaukat Abidi, MaryAnne Williams, and Benjamin Johnston. Human pointing as a robot directive. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 67–68. IEEE, 2013.
- [2] Sergey Adamovich, Michail Berkinblit, Boris Smetanin, Olga Fookson, and Howard Poizner. Influence of movement speed on accuracy of pointing to memorized targets in 3d space. *Neuroscience letters*, 172(1-2):171–174, 1994.
- [3] Simon Alexanderson, Gustav Eje Henter, Taras Kucherenko, and Jonas Beskow. Style-controllable speech-driven gesture synthesis using normalising flows. *Computer Graphics Forum*, 39(2):487–496, 2020.
- [4] Jarrod Blinch, Youngdeok Kim, and Romeo Chua. Trajectory analysis of discrete goal-directed pointing movements: How many trials are needed for reliable data? *Behavior research methods*, 50(5):2162–2172, 2018.
- [5] Luigia Camaioni, Paola Perucchini, Francesca Bellagamba, and Cristina Colonnese. The role of declarative pointing in developing a theory of mind. *Infancy*, 5:291–308, 2004.
- [6] Herbert H. Clark. Coordinating with each other in a material world. *Discourse Studies*, 7(4-5):507–525, 2005.
- [7] Iwan de Kok, Julian Hough, David Schlangen, and Stefan Kopp. Deictic gestures in coaching interactions. In *Proceedings of the Workshop on Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*, pages 10–14, 2016.
- [8] David Droschel, Jörg Stückler, and Sven Behnke. Learning to interpret pointing gestures with a time-of-flight camera. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 481–488, 2011.
- [9] Markus Häring, Jessica Eichberg, and Elisabeth André. Studies on grounding with gaze and pointing gestures in human-robot-interaction. In *International Conference on Social Robotics*, pages 378–387. Springer, 2012.
- [10] Leonard Hasenclever, Fabio Pardo, Raia Hadsell, Nicolas Heess, and Josh Merel. CoMic: Complementary task learning amp; mimicry for reusable skills. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 4105–4115. PMLR, 13–18 Jul 2020.
- [11] Oliver Herbst and Wilfried Kunde. How to point and to interpret pointing gestures? instructions can reduce pointer–observer misunderstandings. *Psychological Research*, 82:395–406, 2018.
- [12] Rachel M Holladay, Anca D Dragan, and Siddhartha S Srinivasa. Legible robot pointing. In *The 23rd IEEE International Symposium on robot and human interactive communication*, pages 217–223. IEEE, 2014.
- [13] Takamasa Iio, Satoro Satake, Takayuki Kanda, Kotaro Hayashi, Florent Ferreri, and Norihiro Hagita. Human-like guide robot that proactively explains exhibits. *International Journal of Social Robotics*, pages 1–18, 2019.
- [14] Sotaro Kita. *Pointing: Where language, culture, and cognition meet*. Psychology Press, 2003.
- [15] Jens Kober, J. Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32:1238–1274, 09 2013.
- [16] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R Thórisson, and Hannes Vilhjálmsson. Towards a common framework for multimodal generation: The behavior markup language. In *International workshop on intelligent virtual agents*, pages 205–217. Springer, 2006.
- [17] Kai Nickel and Rainer Stiefelhagen. Visual recognition of pointing gestures for human–robot interaction. *Image and Vision Computing*, 25(12):1875–1884, 2007. The age of human computer interaction.
- [18] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics (TOG)*, 37(4):1–14, 2018.
- [19] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *arXiv e-prints*, pages arXiv–2104, 2021.
- [20] Camilo Perez Quintero, Romeo Tatsambon, Mona Gridseth, and Martin Jägersand. Visual pointing gestures for bi-directional human robot interaction in a pick-and-place task. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 349–354, 2015.
- [21] Michal Tölgessy, Martin Dekan, František Duchoň, Jozef Rodina, Peter Hubinský, and Luboš Chovanec. Foundations of visual linear human–robot interaction via pointing gesture navigation. *International Journal of Social Robotics*, 9(4):509–523, 2017.
- [22] Herwin Van Welbergen, Ramin Yaghoubzadeh, and Stefan Kopp. Asprealizer 2.0: The next steps in fluent behavior realization for ecas. In *International Conference on Intelligent Virtual Agents*, pages 449–462. Springer, 2014.
- [23] Youngwoo Yoon, Bok Cha, Joo-Haeng Lee, Minsu Jang, Jaeyeon Lee, Jaehong Kim, and Geehyuk Lee. Speech gesture generation from the trimodal context of text, audio, and speaker identity. *ACM Transactions on Graphics (TOG)*, 39(6):1–16, 2020.
- [24] Xuan Zhao, Tingxiang Fan, Dawei Wang, Zhe Hu, Tao Han, and Jia Pan. An actor-critic approach for legible robot motion planner. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5949–5955. IEEE, 2020.