Supplementary Material: Quantifying the Effect of Feedback Frequency in Interactive Reinforcement Learning Performance for Robotic Tasks

Daniel Harnack[1], Julie Pivin-Bachler[2] and Nicolás Navarro-Guerrero[1*]

[1*]Robotics Innovation Center, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH, Robert-Hooke-Straße 1, Bremen, 28359, Bremen, Germany.
[2]Robotics and Interactive Systems – UPSSITECH, University Paul Sabatier, 118 Route de Narbonne, Toulouse, 31062, Occitanie, France.

*Corresponding author(s). E-mail(s): nicolas.navarro@dfki.de;
Contributing authors: daniel.harnack@dfki.de; julie.pivin-bachler@univ-tlse3.fr;

1 NAO 2 DoFs condition

Table 1 summarizes the hyperparameters for the NAO 2 DoFs condition.

Table 1 The hyperparameters used for the NAO 2 DoFs condition.

Hyperparameter	Value
Actor learning rate	$10^{-3.13222995118297}$
Critic learning rate	$10^{-2.3135087887161}$
Exploration rate	0.476312959702095
Discount factor	0.778296920609735
Zeta	−2.18754676532747
Initial variance	1.47012001149771
Actor: # of hidden layers	2
Actor: # of neurons on 1st layer	40
Actor: # of Neurons on 2nd layer	70
Actor: Activation function	ReLu
Critic: # of hidden layers	3
Critic: # of neurons on 1st layer	50
Critic: # of Neurons on 2nd layer	35
Critic: # of Neurons on 3rd layer	35
Critic: Activation function	Softplus

Figure 1 shows the distance distribution of the starting position and goals of the datasets used for the NAO 2 DoFs condition.

2 NAO 4 DoFs condition

Table 2 summarizes the hyperparameters for the NAO 4 DoFs condition.

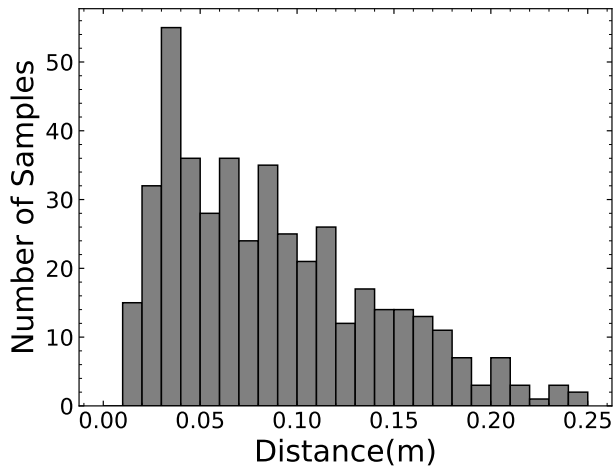

Fig. 1 Distance distribution of the data set used for the NAO 2 DoFs condition.

Table 2 The hyperparameters used for the NAO 4 DoFs condition.

Hyperparameter	Value
Actor learning rate	$10^{-3.56903673957811}$
Critic learning rate	$10^{-2.47908385912353}$
Exploration rate	0.270347209582103
Discount factor	0.808715369645239
Zeta	−1.56020796043399
Initial variance	2.03981318955125
Actor: # of hidden layers	2
Actor: # of neurons on 1st layer	60
Actor: # of Neurons on 2nd layer	75
Actor: Activation function	ReLu
Critic: # of hidden layers	1
Critic: # of neurons on 1st layer	30
Critic: Activation function	Softplus

Figure 2 shows the distance distribution of the starting position and goals of the datasets used for the NAO 4 DoFs condition.

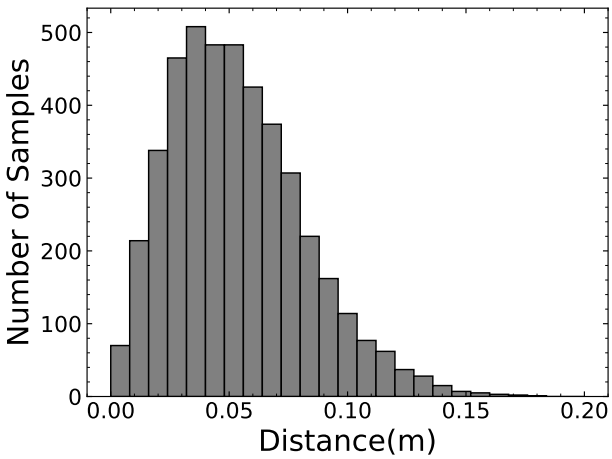

Fig. 2 Distance distribution of the data set used for the NAO 4 DoFs condition.

Figure 3 shows the evolution of the Positioning Error, Positioning Speed, and Failure Rate for the NAO 4 DoFs experiment shown in log scale. The dashed line in the Positioning Error subfigure indicates when the agents reached the Goal Zone Radius (GZR). The horizontal dashed lines in the Positioning Speed subfigure indicate the range of speed possible. The circle indicates the best performance for the corresponding $\mathcal{L}$ value.

3 KUKA 2 DoFs condition

Table 3 summarizes the hyperparameters for the KUKA 2 DoFs condition.

Table 3 The hyperparameters used for the KUKA 2 DoFs condition.

Hyperparameter	Value
Actor learning rate	$10^{-3.26970612771937}$
Critic learning rate	$10^{-3.40694296363931}$
Exploration rate	0.728667095687655
Discount factor	0.961283498715382
Zeta	−1.39245916474946
Initial variance	2.26159240118993
Actor: # of hidden layers	2
Actor: # of neurons on 1st layer	40
Actor: # of Neurons on 2nd layer	50
Actor: Activation function	ReLu
Critic: # of hidden layers	1
Critic: # of neurons on 1st layer	50
Critic: Activation function	ReLu

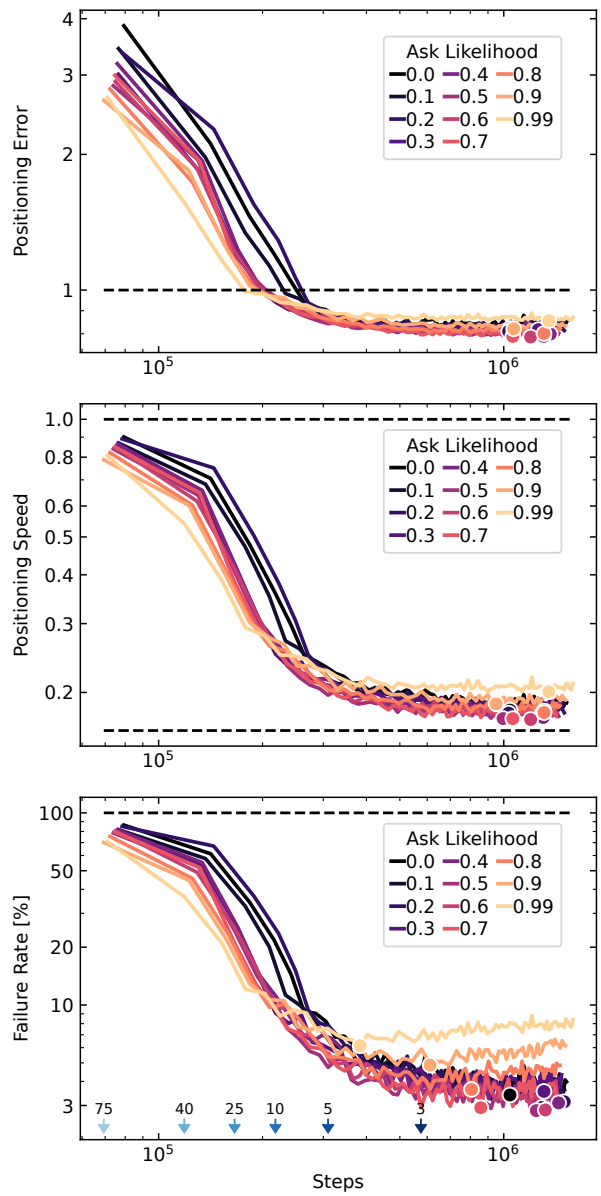

Fig. 3 Evolution of the Positioning Error, Positioning Speed and Failure Rate for the NAO 4 DoFs experiment shown in log scale. The circle indicates the best performance for the corresponding $\mathcal{L}$ value. The blue arrows show the number of environment steps needed for the fastest $\mathcal{L}$ agents to reach 75, 50, 25, 10 and 5% failure rate.

Figure 4 shows the distance distribution of the starting position and goals of the datasets used for the KUKA 2 DoFs condition.

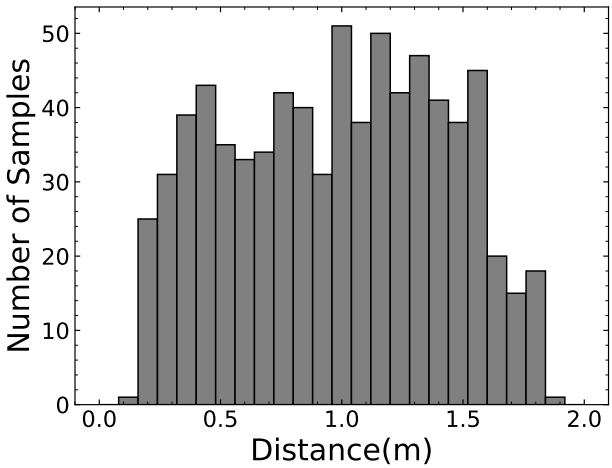

Fig. 4 Distance distribution of the data set used for the KUKA 2 DoFs condition.

4 KUKA 4 DoFs condition

Table 4 summarizes the hyperparameters for the KUKA 4 DoFs condition.

Table 4 The hyperparameters used for the KUKA 4 DoFs condition.

Hyperparameter	Value
Actor learning rate	$10^{-3.7412609002014}$
Critic learning rate	$10^{-4.06033371982199}$
Exploration rate	0.401488943238726
Discount factor	0.97740103054431
Zeta	−2.63270971445589
Initial variance	1.49692672003067
Actor: # of hidden layers	3
Actor: # of neurons on 1st layer	120
Actor: # of Neurons on 2nd layer	90
Actor: # of Neurons on 3rd layer	60
Actor: Activation function	ReLu
Critic: # of hidden layers	3
Critic: # of neurons on 1st layer	100
Critic: # of Neurons on 2nd layer	90
Critic: # of Neurons on 3rd layer	80
Critic: Activation function	ReLu

Figure 5 shows the distance distribution of the starting position and goals of the datasets used for the KUKA 4 DoFs condition.

5 KUKA 7 DoFs condition

Table 5 summarizes the hyperparameters for the KUKA 7 DoFs condition.

Figure 6 shows the distance distribution of the starting position and goals of the datasets used for the KUKA 7 DoFs condition.

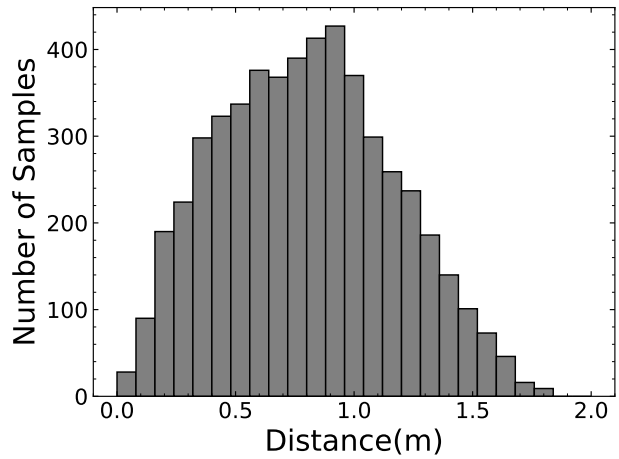

Fig. 5 Distance distribution of the data set used for the KUKA 4 DoFs condition.

Table 5 The hyperparameters used for the KUKA 7 DoFs condition.

Hyperparameter	Value
Actor learning rate	$10^{-4.12762141449049}$
Critic learning rate	$10^{-3.00341554364397}$
Exploration rate	0.343936495522709
Discount factor	0.835172748363745
Zeta	−1.93218433426453
Initial variance	1.49663771483118
Actor: # of hidden layers	3
Actor: # of neurons on 1st layer	100
Actor: # of Neurons on 2nd layer	95
Actor: # of Neurons on 3rd layer	30
Actor: Activation function	ReLu
Critic: # of hidden layers	1
Critic: # of neurons on 1st layer	60
Critic: Activation function	Softplus

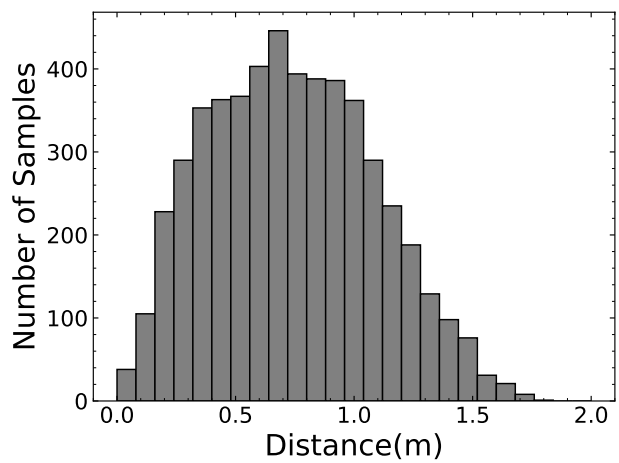

Fig. 6 Distance distribution of the data set used for the KUKA 7 DoFs condition.

Figure 7 shows the evolution of the Positioning Error, Positioning Speed, and Failure Rate for the KUKA 7 DoFs experiment shown in log scale. The dashed line in the Positioning Error subfigure indicates when the agents reached the Goal Zone Radius (GZR). The horizontal dashed lines in the Positioning Speed subfigure indicate the range of speed possible. The circle indicates the best performance for the corresponding $\mathcal{L}$ value.

Funding. Open Access funding provided by the Projekt DEAL (Open access agreement for Germany). Research funding by the *M-RoCK – Human-Machine Interaction Modeling for Continuous Improvement of Robot Behavior* project funded by the Federal Ministry of Education and Research with grant no. 01IW21002.

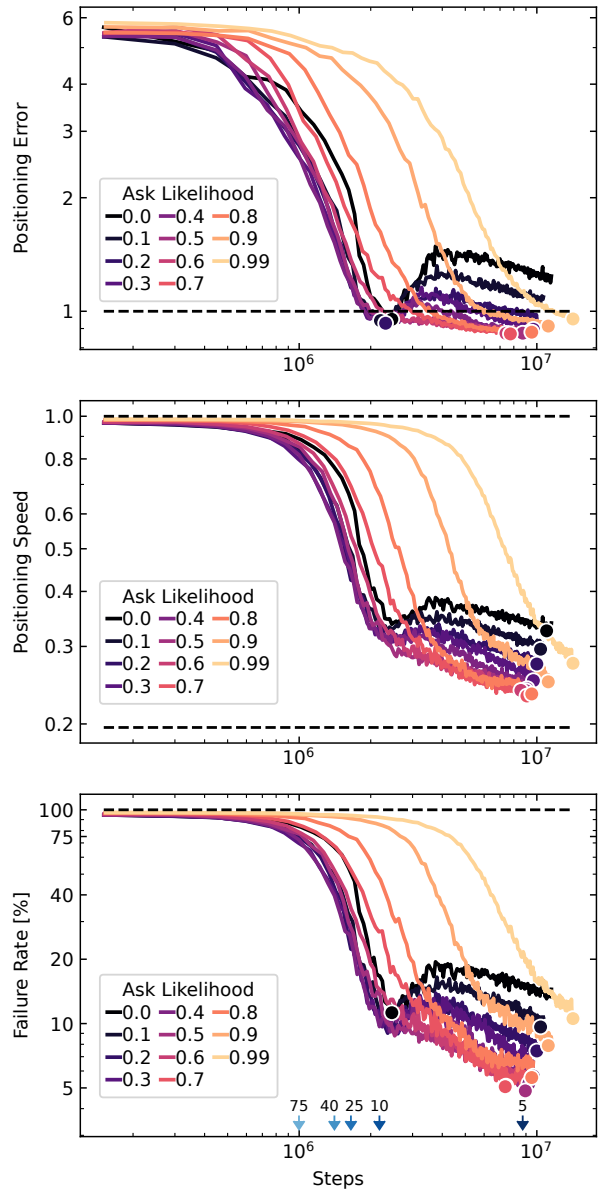

Fig. 7 Evolution of the Positioning Error, Positioning Speed and Failure Rate for the KUKA 7 DoFs experiment shown in log scale. The circle indicates the best performance for the corresponding $\mathcal{L}$ value. The blue arrows show the number of environment steps needed for the fastest $\mathcal{L}$ agents to reach 75, 50, 25, 10 and 5% failure rate.