# Design and Experimental Learning of Swimming Gaits for a Magnetic, Modular, Undulatory Robot

Hankun Deng[1], Patrick Burke[1], Donghao Li[1] and Bo Cheng[1], *Member, IEEE*

***Abstract*— Here we developed an experimental platform with a magnetic, modular, undulatory robot (μBot) for studying fish-inspired underwater locomotion. This platform will enable us to systematically explore the relationship between body morphology, swimming gaits, and swimming performance via reinforcement learning methods. The μBot was designed to be easily modifiable in morphology, compact in size, easy to be controlled and inexpensive. The experimental platform also included a towing tank and a motion tracking system for real-time measurement of the μBot kinematics. The swimming gaits of μBot were generated by a central pattern generator (CPG), which outputs voltage signals to μBot's magnetic actuators. The CPG parameters were learned experimentally using the parameter exploring policy gradient (PGPE) method to maximize swimming speed. In the experiments, two μBot designs with the same body morphology but different caudal-fin shapes were tested. Results showed that swimming gaits with back-propagating traveling waves can be learned experimentally via PGPE, while the shape of the caudal fins had moderate influences on the learned gaits and the swimming speed. Furthermore, robot swimming speed was sensitive to the undulating frequency and the voltage magnitude of the last three posterior actuators. In contrast, swimming gaits and speed were relatively invariant to the variances within the inter-module connection weights of CPG and the voltage applied to the anterior actuator.**

## I. Introduction

Fish swimming is the epitome of successful and diverse forms of underwater locomotion, which spans a wide range of body size and speed [1], [2], [3], and a large spectrum of Reynolds number and Strouhal number [4], [5]. It is hardly surprising that engineers often draw inspirations from the morphologies and kinematics of fish swimming for novel underwater propulsion [6], [7], [8], [9]. However, it is also remarkably challenging to model and emulate fish swimming for robotics, especially due to its large morphological and kinematic design space and the difficulties in understanding the relationship between its diverse forms and functions.

Fish locomotion can be characterized by their propulsion mechanisms, which are generally categorized into two forms: Body and/or Caudal Fin (BCF) and Median and/or Paired Fin (MPF). Most fish species in nature use BCF for swimming (approximately 85%, [3]). While BCF forms achieve higher speed, MPF forms offer better maneuverability [4]. Fish locomotion is usually investigated using one or a combination of the following three methods: experiments with biological fish [10], [11], experiments or simulation of robotic fish [12], [13], or computational fluid dynamics (CFD) simulation of biological fish [14], [15]. Biological fish experiments can directly reveal the biomechanics of a particular fish swimming behavior under investigation, such as backward swimming and vortex exploitation, however it does not allow systematic and large-scale investigation on fish morphologies and swimming gaits, as well as their relationships. CFD is a powerful tool that allows for modeling and modulation of fish morphologies and gaits and provides high-fidelity fluid flow and pressure data for swimming physics. However, the high computational cost makes it still impractical for systematic investigation on fish morphologies and swimming gaits, especially using optimization or learning methods [16] (with one recent exception, [17]).

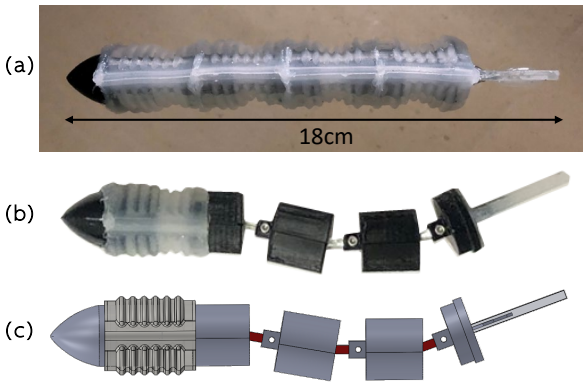


Fig 1. Overview of the assembled μBot. (a) Top view of a μBot in water with outer suits on. (b) Top view of a μBot with outer suits removed except the first one. (c) Model of a μBot with outer suits removed except the first one.

Although design and construction of robotic fish is no easy task, they can be fully manipulated in forms and controlled in gaits, and therefore provide an excellent platform for not only studying biological fish swimming (i.e., robotics-inspired biology, [18], [19]) but also for investigations of fish-inspired underwater locomotion methods in general. In fact, recent decades have seen plenty of successful robotic fish designs which were used as platforms to study the underlying mechanisms of robot-fluid interaction [6], [13], test control strategies [7], [20] or explore underwater environments [8], covering four main BCF swimming modes (anguilliform, subcarangiform, carangiform and thunniform), although there are still unquestionably large gaps in swimming performance between the biological and robotic fish.

Notably, with the progress on robotic fish designs, systematic investigations on fish form and function relationships, including experimental learning of swimming gaits for novel fish-inspired underwater propulsion are still scarce in the literature. In this work, we developed an experimental platform with a magnetic, modular undulatory robot ($M^2$UBot, or μBot) for systematic investigations of the

Research supported by National Science Foundation (CNS-1932130 awarded to B.C) and Army Research Office (W911NF-20-1-0226, awarded to B.C).

[1]Department of Mechanical Engineering, Penn State University, University Park, PA, 16802, USA. hxd202@psu.edu, buc10@psu.edu.

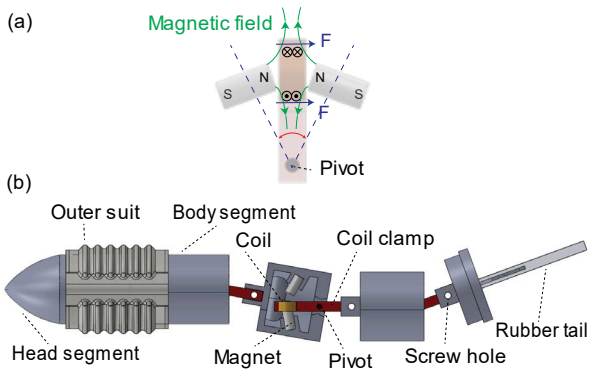


Fig 2. Design of the whole robot and the magnetic actuator. (a) Operating principles of the actuator, (b) Top-view of a μBot with 4 actuators (outer suit is removed except the first one for illustration purposes). The third segment shows the cross-section view of the actuator.

relationships among the body morphologies, swimming gaits, and performance. With the modular design of body segments (based on magnetic actuators) and its easy construction and assembly into the full robot, body morphologies of the μBot can be relatively easily modified. In addition, the swimming gaits of the robot were generated by a central pattern generator (CPG), the parameters of which can be learned by a policy search reinforcement learning (RL) algorithm to maximize the swimming speed, using an experimental platform including a towing tank, motion tracking system and μBot itself.

In this work, for two different caudal-fin shape designs of μBot, we successfully used this platform to optimize the gaits for steady swimming speed, thereby investigated the relationship among caudal-fin morphology, optimized swimming gaits and speed. The rest of this paper is organized as follows. In section II, the design and construction of μBot are introduced. Section III explains the CPG design of μBot. Section IV presents the gait learning problem and the learning algorithm. Then the experimental setup is described in Section V and the learning results are presented in section VI. Finally, discussions and future work are summarized in section VII.

## II. Design of μBot

Here we describe the design aspects and assembly process of μBot. To systematically investigate the form and function relationships for underwater locomotion, we aimed at creating a robot which is easily modifiable in design, compact in size, easy to control and inexpensive. The compact size of the robot allows the experiments to be conducted in more controlled lab settings with limited space, while relatively low cost enables easy prototyping and testing with more prototypes.

### A. Magnetic, modular actuator

Several types of actuators have been used for swimming robots, such as electric motors [7], [13], [20], hydraulic/pneumatic actuators [8], [21], magnetic actuators [22], [23], [24] and smart material-based actuators [25]. Among these actuators, electric motors are most widely used because of its low cost, reliable performance, and high efficiency. However, it is still not easy for motors to scale down because of their relatively complex inner structure. Here we chose to use a magnetic actuator that has a simple design (Fig. 2a and b) and can be easily modularized and scaled up or down for body segments of various sizes (similar to biological fish). For the two μBot designs presented in this study, all segments have a uniform size along the body.

The magnetic actuator has a coil mounted on a rotating arm (coil clamp) around a pivot joint, while the coil is placed in between two permanent magnets pointed closely to each other with the same polarity (therefore opposing each other, Fig. 2a and b). By applying voltage to the coil, it generates a magnetic field approximately orthogonal to those generated by the permanent magnets (nearly radially symmetric), which results in an electromagnetic force that rotates the coil, coil-clamp, and the next segment around the pivot. Reversing the current/voltage will simply reverse the rotation direction, and periodic voltage input will generate oscillatory rotational motion. The range of the rotation angle of this actuator is $\pm 20^{o}$ to balance the mobility of the actuator and the magnitude of the generated force.

### B. Robot outer soft suit

The gaps between the modular body segments are covered by soft rubber suits, which provide a continuous surface for robot-fluid interaction, the morphology of which can be easily varied for different μBot designs. We used a uniform design of the rubber suit for all segments (Fig. 2b). The rubber suit is made of silicone rubber (Ecoflex 00-30, Smooth-On Inc, PA, USA), and it also helps to waterproof the robot and provides body compliance for better fluid-structure interaction. The crinkled shape on the rubber suit surface is to reduce the compliance but can also be easily varied to modify compliance, although we did not investigate the influence of body stiffness on the swimming performance in this work.

### C. Robot Assembly

The μBot designs used in this work for gait-learning experiments have 5 segments with 4 actuators, including the head and caudal-fin segments. The assembled robot is shown in Fig. 1. Since the coil generates heat during operating, the coil clamp is made of Aluminum 6061-T6 using CNC machining. The internal frame of μBot is 3D printed with PETG material. Each rubber suit segment is glued together by sil-poxy with the adjacent ones. The total length of μBot is 18 cm. The depth is 4 cm and the width in lateral direction is 2.8 cm. The total weight of the robot is around 82 g. The robot is designed so that its average density is slightly less than water and buoyancy equilibrium is achieved with a minor part of the body above the surface (10% in depth), which is common in swimming robots like AmphiBot [7] and salamander robot [26]. During experiments, reflective markers are attached on top of the robot for camera detection. In testing, the robot shows excellent durability, and can last more than one thousand tests, with each test taking 10s.

## III. CPG Design for μBot Swimming

The swimming gaits of the μBot are generated by a CPG network, which outputs rhythmic voltage signals to the magnetic actuators. The most basic feature of CPGs is to map non-patterned, low-dimensional control commands to well-coordinated, high-dimensional rhythmic motor inputs [27]. Since most bio-inspired robotic locomotion requires rhythmic motion patterns, CPGs have been widely applied to bipedal

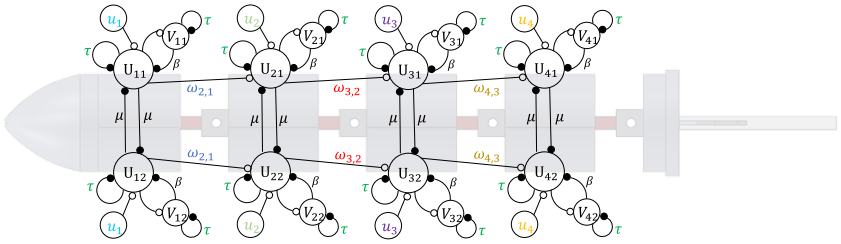

Fig 3. The complete CPG network of μBot with its parameters labeled for each actuator. Parameters labeled with colors were learned experimentally, while those in black were fixed. The subscripts of the parameters were simplified according to how they were learned, see TABLE I.

robots [28], [29], crawling robots [26], [30], and swimming robots [31], [32].

Among the different mathematical models of CPGs, the model proposed by Matsuoka [33] is adopted here for μBot. Specifically, we used a dual-neuron model for each actuator, i.e., each actuator contains two neurons that inhibit each other. The mathematical expression of each neuron is represented by two ordinary differential equations (ODEs) as follows,

$$\tau_{1,i,j}\dot{U}_{i,j} + U_{i,j} = u_{i,j} - \beta_{i,j}V_{i,j} - \mu_{i,j}y_{i,3-j} + \sum_{k=1}^{n}\omega_{i,k}y_{k,j}$$

$$\tau_{2,i,j}\dot{V}_{i,j} + V_{i,j} = y_{i,j}$$

$$y_{i,j} = \max(0, U_{i,j}) \quad (1)$$

$$y_{i,out} = y_{i,1} - y_{i,2}$$

$$i,k = 1,2,\dots,n, i \neq k; j = 1 \text{ or } 2$$

where $n$ is CPG module number; $\tau$ is the time constant for each neuron; $u$ represents external stimulus for each neuron; $\beta$ and $\mu$ are adaption coefficient and mutual inhibition weight in one module; $\omega$ is the inter-module connection weight of the neuron; $y_{i,out}$ is the output of the $i^{th}$ CPG module [30].

Generally, rhythmic signals generated from modular CPGs can be characterized by their individual magnitude and frequency and their inter-module phase delay. However, there exists certain parameter redundancy in the Matsuoka CPG model for our specific application. Therefore, to reduce the number of parameters for experimental gait-learning, some parameters were fixed while others were learned (TABLE I). Specifically, the following CPG parameters were learned: $\alpha = [\tau, \omega_{2,1}, \omega_{3,2}, \omega_{4,3}, u_1, u_2, u_3, u_4]^{\mathrm{T}}$. Fig. 3 illustrates the complete CPG structure with all parameters labeled.

## IV. Swimming Gait Learning – Policy Search

### A. Gait-learning problem

Underwater locomotion and swimming performance contain many detailed aspects, such as speed, efficiency, maneuverability, and stability. Here we chose to optimize the μBot's gaits for swimming speed. Notably, the swimming gaits and speed are inseparable, emergent behaviors that arise from the interactions between the fluids and μBot's body deformable structure, controlled by CPG-generated actuator voltage inputs and also dependent on the body morphologies, such as the tail shape. Li et al. have shown with numerical simulation that forked-shape tail can increase both mean thrust and efficiency, compared with the rectangular ones [34]. Experiments on Tunabot also show that the swimming speed can be largely influenced by tail beat frequency [13]. Gazzola et al. have built a mathematical model to illustrate how the swimming speed can be influenced by varying body wave forms [35]. In this work, our goal is to maximize the steady forward swimming speed. During testing, μBot took less than 6s for acceleration in general. Therefore, we set the μBot to swimming forward for 10s and used the average speed within the last 3 seconds as the reward for the learning.

As discussed in section III, the learned parameters vector is $\alpha = [\tau, \omega_{2,1}, \omega_{3,2}, \omega_{4,3}, u_1, u_2, u_3, u_4]^{\mathrm{T}}$. However, $\omega$ and $u$ are not contributing to the model individually. Instead, $\omega/\tau$ and $u/\tau$ are the terms that influence the states of the ODEs. Thus, $\omega$ and $u$ are normalized by $\tau$. Also, the parameters are scaled (see below) so that a single learning rate can be applied to all parameters during learning. Consequently, the parameters vector to be learned is:

$$v = \alpha^* = [\tau^*, \omega_1^*, \omega_2^*, \omega_3^*, u_1^*, u_2^*, u_3^*, u_4^*]^{\mathrm{T}}$$
$$= [100\tau, 100\frac{\omega_{2,1}}{\tau}, 100\frac{\omega_{3,2}}{\tau}, 100\frac{\omega_{4,3}}{\tau}, \frac{u_1}{\tau}, \frac{u_2}{\tau}, \frac{u_3}{\tau}, \frac{u_4}{\tau}]^{\mathrm{T}}.$$

TABLE I. CPG PARAMETERS

| Parameter | Learned or fixed | Value or comments |
|---|---|---|
| $\boldsymbol{\tau_{1,i,j}}$ | Learned | same for all neurons |
| $\boldsymbol{\tau_{2,i,j}}$ | Learned | same with $\boldsymbol{\tau_{1,i,j}}$ |
| $\boldsymbol{u_{i,j}}$ | Learned | same in one module, different between modules |
| $\boldsymbol{\beta_{i,j}}$ | Fixed | 4.5 for all the neurons |
| $\boldsymbol{\mu_{i,j}}$ | Fixed | 3 for all the neurons |
| $\boldsymbol{\omega_{i,k}}$ | Learned | $\boldsymbol{\omega_{i,k}} \neq \mathbf{0}$ only if $\boldsymbol{k} = \boldsymbol{i} - \mathbf{1}$, same in one module, different between modules |

### B. Parameter Exploring Policy Gradient

Regarding the training of CPGs, different reinforcement learning algorithms have been applied, like actor-critic method [36] and parameters exploring policy gradient (PGPE) [37]. In this work, we applied PGPE method for the experimental μBot gait learning. PGPE is deterministic within each rollout, as the entire rollout history is generated from the parameters sampled from probabilistic policy parameter distributions, so that the variance in the gradient estimation is reduced [38], [39].

Specifically, in PGPE, the policy parameters $v$ are sampled from the probability distribution $p(v|\rho)$ where $\rho$ is the hyperparameters vector governing the sampling of the policy parameters. Based on the returned reward $R(v)$, the hyperparameters $\rho$ can be updated as,

$$\rho \leftarrow \rho + \gamma\nabla_{\rho}J(\rho), \quad (2)$$

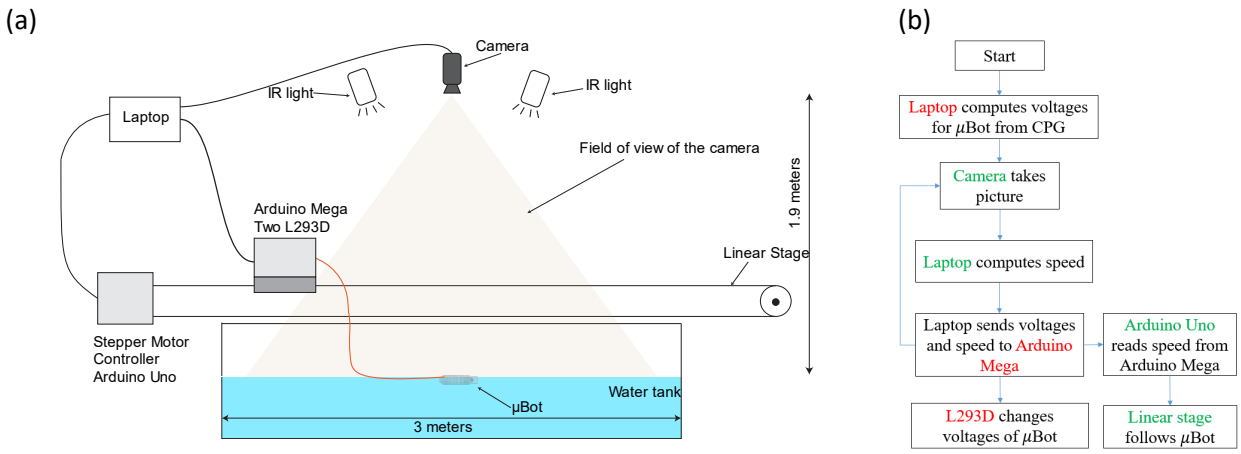


Fig 4. (a) Experimental setup of μBot learning. (b) Operating procedure of the whole system: red color marks the key elements for control system and green color marks the key elements for motion tracking system.

where $\gamma$ is the learning rate. The gradient of the value function $\nabla_\rho J(\rho)$ is derived as ([38]),

$$\nabla_\rho J(\rho) \approx \frac{1}{N}\sum_{i=1}^{N} \nabla_\rho \log p(v^i|\rho) r(v^i), \qquad (3)$$

where $N$ represents the number of rollouts and $r(v^i)$ denotes the reward for $i^{th}$ sampled parameters vector. The hyperparameters $\rho = [\mu\ \sigma]^T$ include $\mu$ and $\sigma$ for the mean and standard deviation of the normal distribution of CPG parameters vector $v$. To ensure a positive value of the standard deviation, the lower bound of $\sigma$ is set as $10^{-4}$ for all the 8 CPG parameters.

Baseline helps to decrease the variance of a policy gradient [40], allowing faster learning. In terms of the baseline selection, there are many options. In this work, the idea of optimal baseline in [41] is applied and the equation of the baseline is expressed as:

$$b = \sum_{i=1}^{N} r(v^i) \left\|\nabla_\rho \log p(v^i|\rho)\right\|^2 / \sum_{i=1}^{N} \left\|\nabla_\rho \log p(v^i|\rho)\right\|^2. \quad (4)$$

The parameter vector $v$ contains 8 parameters. Empirically, the learning curve shows good convergence with 15 rollouts. Therefore, we chose $N = 15$ to balance the estimation accuracy and experimental cost.

To reduce the number of episodes and prevent overshoot of $\sigma$ during the experimental learning, we used an empirical, heuristic learning rate adjusting method. First, the nominal value of $\gamma$ was set as 0.5 for the first 6 episodes and 1 for the rest. During the experiments, the standard deviation of $v$ might decrease to the lower bound very quickly and stop exploration further. We prevented this situation within the first 6 episodes, by setting the learning rate as 0.2 if any $\sigma$ decreases below the lower bound. In addition, when the normalized variance (variance/mean) of the returned reward $r(v)$ was less than 0.2 (the update of $\rho$ will be small), the learning rate was set as 1 rather than 0.5. The learning was stopped when the normalized variance of the returned reward $r(v)$ was less than 0.03, which indicated a near locally optimal solution has been identified.

## V. Swimming Gait Learning – Experimental Setup

The experimental platform contains two subsystems: a real-time control system that outputs CPG signals to the actuators, and a motion tracking system that measures the μBot's gaits and speed. The sketched experimental setup is illustrated in Fig. 4a.

The control system consists of a laptop, a microcontroller (Arduino Mega), two L293D chips and an external DC power supply. The laptop generates the CPG signals, which are first sent to the microcontroller through serial communication, and then to the L293D to drive the actuators. The voltage from the DC power supply was set as 12V to protect the coil of μBot.

The motion tracking system consists of the same laptop as used in the control system, a monochrome camera (acA2000-165umNIR, Basler AG Inc, Ahrensburg, Germany) and a 760nm filter, IR light sources, a microcontroller (Arduino Uno) and a linear stage. During the experimental learning, the head position of μBot was captured by the camera and sent to the laptop at each time step. Then the forward swimming speed was calculated using backward difference. The μBot speed signal, after being filtered, was also sent to the linear stage through serial communication (Fig. 4b). The linear stage, which carried the wires of the μBot, was moved at the same speed of μBot to make the μBot wires tension free, thereby removing its effect on the swimming performance. The whole system operated at 20Hz. The robot was brought back to the original position after each trail.

## VI. Experimental Results

In the current work, μBot designs with two tail shapes were tested, one with a rectangular shape (μBot-1) and the other with an inclined bottom edge (μBot-2) (Fig. 5). The rectangular tail had a similar shape with that of AmphiBot III in [7] and the inclined tail took the shape $E_2^{30}$ used in [39], which had the best thrust-generation efficiency among those tested. For each μBot, the learning processes were repeated three times with different initial conditions to investigate whether the same local optimum can be found. Due to the uncertainty in PGPE sampling, the number of episodes for each learning process is different. On average, a complete learning experiment required approximately 15 episodes to converge. The learning plots for the reward and the CPG parameters are shown in Fig. 5 and 6, respectively.

Since both the control and the motion tracking systems operated at 20Hz and were not phase-locked with the swimming gaits, the sampling of the CPG signal and

swimming gait kinematics can be 5 or 6 per cycle. To have a better representation and visualization of the CPG inputs and swimming gaits, the respective data within the last 3 seconds were assembled into one cycle according to their phases. Fig. 7 shows the CPG voltage signals and swimming gait variables from the last learning episode of each experiment, while the swimming gaits were smoothed out.

### *A. μBot with the rectangular tail (μBot-1)*

Fig. 5a illustrates the reward curves for μBot-1. All the three experiments yielded similar final rewards at approximately 25mm/s as well as similar CPG input signals (with slight differences in phases, Fig. 7a) and swimming gaits (Fig. 7d), which indicated that all three experiments converged closely to a locally optimal gait.

Fig. 7g shows an illustration of the learned gaits for μBot-1. Joint-1 reaches the peak first (at 13.04% of the cycle), while other joints, including the passive tail joint, reach the peaks sequentially in order (dashed line). Notably, even though the joint amplitudes were small, there was a backward traveling wave, propagating from head to tail (same in Fig. 7d). The bending amplitudes of all actuated μBot joints in learned gaits were within $5^o$, which was significantly smaller than the joint/actuator limit ($20^o$). However, the passive bending angle of the tail trailing-edge (red dot, second to the last dot) can have more than $25^o$ (Fig. 7d). A closer examination shows that the phase of the passive bending of tail trailing edge was slightly ahead of the phase of joint-1. Since the distance from the joint-1 (light blue, the second dot) to the passive joint of tail trailing edge (red, second to the last dot) was less than a body length, the whole μBot body consisted of a wavelength larger than one. Remarkably, this is consistent with those observed in biological fish [42] and predicted by Lighthill [43].

Regarding the CPG parameters, $\tau^*$, which determines the frequency of the learned gaits, had the fastest convergence and smallest differences among the three learning experiments (Fig. 6a). The $\omega^*$ terms, which mainly determines the phase differences between two adjacent actuators, showed large differences between experiments and large variances within each experiment even after the reward converged (Fig. 6b, c, and d). Interestingly, the phase delays of the learned CPG voltage signals, as well as the resulting gaits, showed much smaller differences among experiments (Fig. 7a and d). For the $u^*$ terms, which control the magnitude

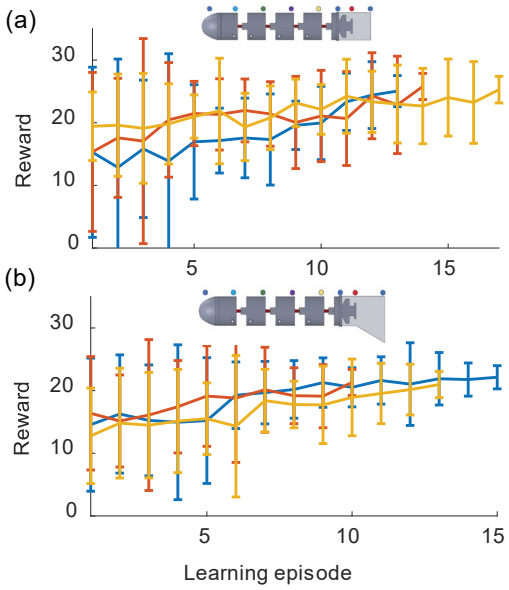


Fig 5. Reward (forward speed in *mm/s*) curves for the two μBots, while the lateral view of μBots with 2 tail designs is shown on top of the reward curves. For each robot, learning is repeated 3 times with varying initial conditions. The vertical bars are the range of mean ± 3*standard deviation. (a) μBot-1. (b) μBot-2.

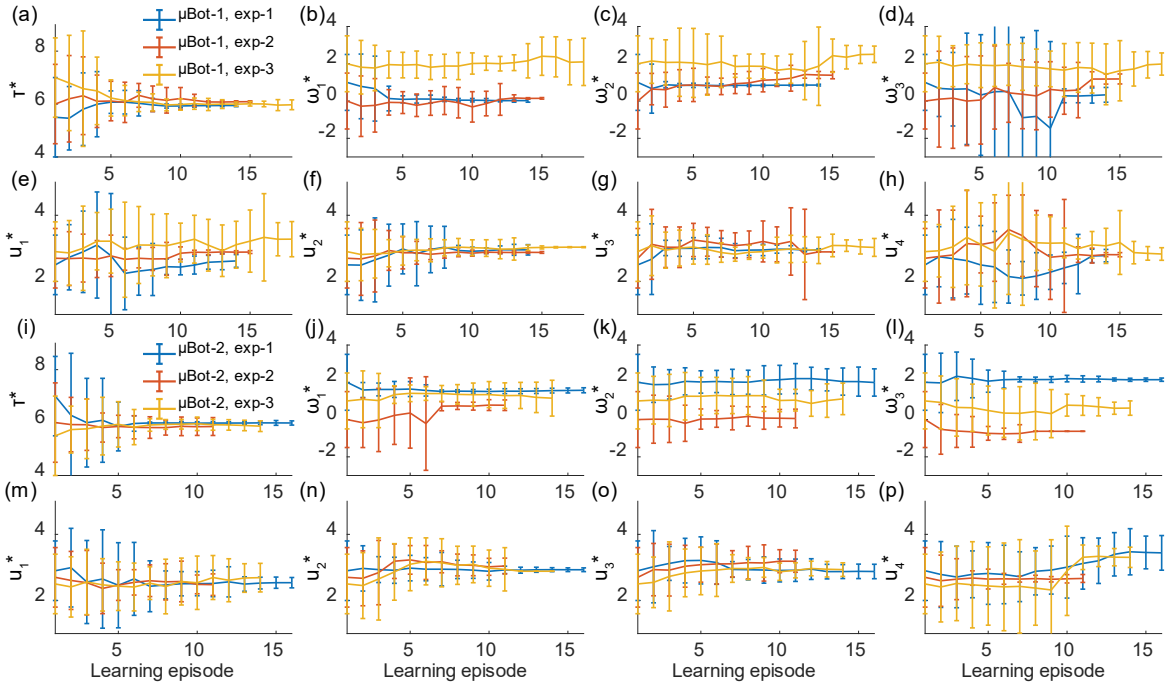


Fig 6. Learning curves of all the learned parameters. For each robot, learning is repeated 3 times with varying initial conditions. Vertical bars are the range of mean ± 3*standard deviation. (a)-(h) are learning plots for μBot-1. (i)-(p) are learning plots for μBot-2.

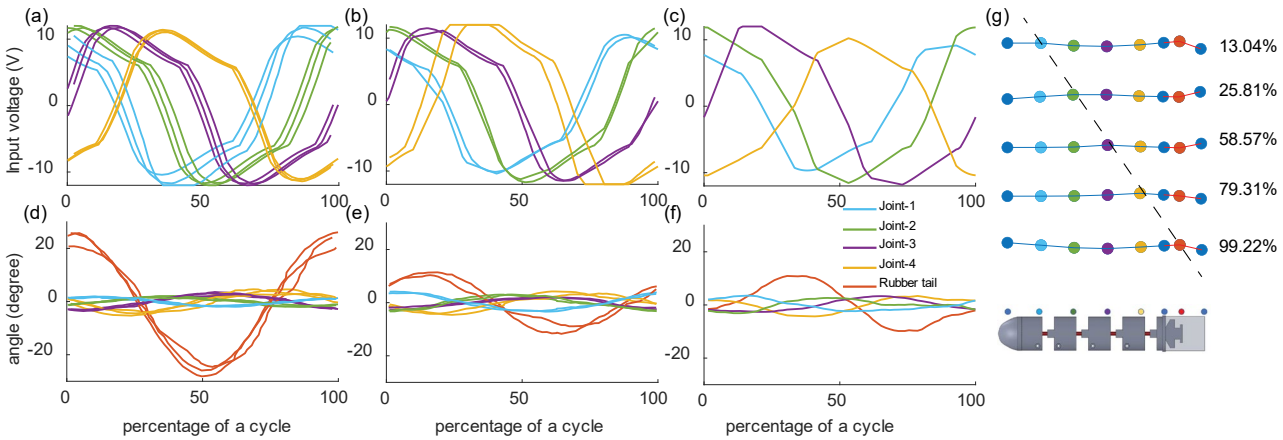


Fig 7. Learned voltage signals and swimming gaits. Joint-1 to joint-4 correspond to the 4 actuators in order from head to tail. (a) (d): μBot-1, exp-1, exp-2, and exp-3. (b) (e): μBot-2, exp-1 and exp-3. (c)(f): μBot-2, exp-2. (g): illustration of a learned gait of μBot-1, where the color dots match the corresponding positions on the robot body. The blue lines represent the body, and the red lines represent the rubber tail. The dashed line goes through the joints that reach the peak bending angle, indicating that the wave is traveling backward along the body. The percentages on the right side indicate the relative phase of a gait cycle.

of the applied voltage, the final values of $u_2^*$, $u_3^*$and $u_4^*$ in the three experiments were close and the variances at the end of the learning were small (Fig. 6f, g, and h). However, the values of $u_1^*$ had relatively large differences between different experiments and larger variance within each experiment (Fig. 6e). It can also be seen from Fig. 7a that the voltage amplitudes of the last three body joints were all near the voltage bound ($12V$) but that for the most anterior joint could drop to approximately $10V$.

### *B. μBot with the inclined tail (μBot-2)*

For the μBot with the inclined tail, the three learning experiments yielded two distinguishable gaits, however with similar rewards at approximately 21mm/s (Fig. 5b). Exp-1 and 3 yielded the first gaits (Fig. 7b and e) and Exp-2 yielded the second gait (Fig. 7c and f). The main difference between the two gaits is that the phase delays of joint-4 (most posterior actuator) actuation and passive tail joint were both noticeable larger in the second gait than the first. In addition, the voltage magnitude of joint-4 (Fig. 7b and c) in the first gait was higher than that of the second gait and showed saturation to the 12V limit. Nonetheless, similar to μBot-1, travelling waves can be observed in all learned gaits for μBot-2, the joint bending amplitude remained at approximately $5^o$. However, the passive bending amplitudes of the tail trailing edge in μBot-2 were only less than half (approximately $10^o$) of those in μBot-1.

CPG parameters in μBot-2 showed similar convergence behaviors as those in μBot-1, except $u_4^*$ where the standard deviation in Exp-1 is still large after the reward converged (Fig. 6p). However, the distribution of the value of $u_4^*$ term is in a range where the voltage can always reach the bound ($12V$) with more or less saturation.

## VII. Discussions and Future Work

Among all CPG parameters, the $\tau^*$ term showed the fastest convergence, the lowest variances within individual experiment and arrived at the closest value for all learning experiments in both μBot designs. Since $\tau^*$ is the only parameter that determines the undulating gait frequency (approximately 3.9Hz), its good convergence property indicates that the swimming speed is highly sensitive to the driving frequency, which is in agreement with a recent computational result in [44]. In contrast, $\omega^*$ terms were more dispersed, and some had large variances even when the reward converged.

Interestingly, the propagation of variances from CPG parameters to CPG voltage outputs, then to swimming gaits and finally to the swimming speed showed a decaying trend. The convergence results of $u^*$ suggest that the swimming speed is sensitive to the voltage amplitudes of the last three posterior actuators but not to the most anterior one.

In addition, it is noticeable that the recoil of the head (head oscillation) was significantly larger at the start of the learning than that of the converged gaits (result not shown). As per authors' visual inspections of the experimental learning process, the recoil problem was gradually reduced as the swimming performance improved. This may indicate that μBot propagating slightly more than one complete traveling wave along its body length may help to reduce the head recoil, while the swimming performance benefits accordingly, which was previously proposed in [43], [45].

Although we were not attempting to optimize the tail shape of μBot, comparing μBot-1 and μBot-2 with different tail designs, our results did suggest that the optimal gaits were dependent on the robot morphology. However, substantial future work will be needed to further reveal how optimal swimming behaviors emerge from the interactions between the fluids and μBot's body deformable structure.

Finally, note that, in the current learning experiments, we were conservative in setting the actuator voltage limit to 12V to prevent overheating the coils, while a more accurate voltage limit for the long-term safe operation of μBot is yet to be determined. As a result, the body undulatory amplitude of the learned μBot gaits was notably small relative to the physical limit of the joints (20°). In future work, we will further optimize the size of the coil to improve the torque generation of the actuators. More importantly, we will continue to use the μBot platform for systematically investigating the relationships among body morphologies, swimming gaits, and swimming performance.

## REFERENCES


[1] R. Bainbridge, "The Speed of Swimming of Fish as Related to Size and to the Frequency and Amplitude of the Tail Beat," *J. Exp. Biol.*, vol. 35, no. 1, pp. 109–133, 1958.

[2] J. J. Videler and C. S. Wardle, "Fish swimming stride by stride: speed limits and endurance," *Rev. Fish Biol. Fish.*, vol. 1, no. 1, pp. 23–40, 1991,

[3] J. J. Videler, *Fish Swimming*. London, U.K: Chapman and Hall, 1993.

[4] M. Sfakiotakis, D. M. Lane, and J. B. C. Davies, "Review of fish swimming modes for aquatic locomotion," *IEEE J. Ocean. Eng.*, vol. 24, no. 2, pp. 237–252, 1999,

[5] M. Gazzola, M. Argentina, and L. Mahadevan, "Scaling macroscopic aquatic locomotion," *Nat. Phys.*, vol. 10, no. 10, pp. 758–761, 2014,

[6] M. S. Triantafyllou and G. S. Triantafyllou, "An efficient swimming machine," *Sci. Am.*, vol. 272, no. 3, pp. 64–70, 1995.

[7] M. Porez, F. Boyer, and A. J. Ijspeert, "Improved lighthill fish swimming model for bio-inspired robots: Modeling, computational aspects and experimental comparisons," *Int. J. Rob. Res.*, vol. 33, no. 10, pp. 1322–1341, 2014,

[8] R. K. Katzschmann, J. DelPreto, R. MacCurdy, and D. Rus, "Exploration of underwater life with an acoustically controlled soft robotic fish," *Sci. Robot.*, vol. 3, no. 16, 2018,

[9] S.-J. Park *et al.*, "Phototactic guidance of a tissue-engineered soft-robotic ray," *Science*, vol. 353, no. 6295, pp. 158–162, 2016,

[10] K. D'Août and P. Aerts, "A kinematic comparison of forward and backward swimming in the eel Anguilla anguilla," *J. Exp. Biol.*, vol. 202, no. 11, pp. 1511–1521, 1999.

[11] J. C. Liao, D. N. Beal, G. V Lauder, and M. S. Triantafyllou, "Fish exploiting vortices decrease muscle activity," *Science*, vol. 302, no. 5650, pp. 1566–1569, 2003.

[12] R. Mason and J. W. Burdick, "Experiments in Carangiform robotic fish locomotion," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2000, vol. 1, pp. 428–435.

[13] J. Zhu, C. White, D. K. Wainwright, V. Di Santo, G. V. Lauder, and H. Bart-Smith, "Tuna robotics: A high-frequency experimental platform exploring the performance space of swimming fishes," *Sci. Robot.*, vol. 4, no. 34, 2019,

[14] I. Borazjani and F. Sotiropoulos, "Numerical investigation of the hydrodynamics of carangiform swimming in the transitional and inertial flow regimes," *J. Exp. Biol.*, vol. 211, no. 10, pp. 1541–1558, 2008,

[15] X. Chang, L. Zhang, and X. He, "Numerical study of the thunniform mode of fish swimming with different Reynolds number and caudal fin shape," *Comput. Fluids*, vol. 68, pp. 54–70, 2012,

[16] Y. E. Bayiz, S. Hsu, A. N. Aguiles, Y. Shade-alexander, and B. Cheng, "Experimental Learning of a Lift-Maximizing Central Pattern Generator for a Flapping Robotic Wing," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2019, pp. 1997–2003.

[17] S. Verma, G. Novati, and P. Koumoutsakos, "Efficient collective swimming by harnessing vortices through deep reinforcement learning," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 23, pp. 5849–5854, 2018,

[18] N. Gravish and G. V. Lauder, "Robotics-inspired biology," *J. Exp. Biol.*, vol. 221, no. 7, 2018,

[19] A. J. Ijspeert, "Biorobotics: Using robots to emulate and investigate agile locomotion," *Science*, vol. 346, no. 6206, pp. 196–203, 2014.

[20] J. Yu, M. Tan, S. Wang, and E. Chen, "Development of a biomimetic robotic fish and its control algorithm," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 34, no. 4, pp. 1798–1810, 2004,

[21] K. Suzumori, S. Endo, T. Kanda, N. Kato, and H. Suzuki, "A bending pneumatic rubber actuator realizing soft-bodied manta swimming robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2007, pp. 4975–4980.

[22] B. Cheng, J. A. Roll, and X. Deng, "Modeling and optimization of an electromagnetic actuator for flapping wing micro air vehicle," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2013, pp. 4035–4041.

[23] J. A. Roll, B. Cheng, and X. Deng, "Design, Fabrication, and Experiments of an Electromagnetic Actuator for Flapping Wing Micro Air Vehicles," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2013, pp. 809–815.

[24] J. A. Roll, B. Cheng, and X. Deng, "An Electromagnetic Actuator for High-Frequency Flapping-Wing Microair Vehicles," *IEEE Trans. Robot.*, vol. 31, no. 2, pp. 400–414, 2015,

[25] B. Kim, D. H. Kim, J. Jung, and J. O. Park, "A biomimetic undulatory tadpole robot using ionic polymer-metal composite actuators," *Smart Mater. Struct.*, vol. 14, no. 6, pp. 1579–1585, 2005,

[26] A. J. Ijspeert, A. Crespi, D. Ryczko, and J. M. Cabelguen, "From swimming to walking with a salamander robot driven by a spinal cord model," *Science*, vol. 315, no. 5817, pp. 1416–1420, 2007,

[27] S. Grillner and A. E. Manira, "Current principles of motor control, with special reference to vertebrate locomotion," *Physiol. Rev.*, vol. 100, no. 1, pp. 271–320, 2020,

[28] G. Taga, Y. Yamaguchi, and H. Shimizu, "Self-organized control of bipedal locomotion by neural oscillators in unpredictable environment," *Biol. Cybern.*, vol. 65, no. 3, pp. 147–159, 1991,

[29] T. Mori, Y. Nakamura, M. A. Sato, and S. Ishii, "Reinforcement learning for a CPG-driven biped robot," *Assoc. Adv. Artif. Intell.*, vol. 4, pp. 623–630, 2004.

[30] X. Wu and S. Ma, "CPG-based control of serpentine locomotion of a snake-like robot," *Mechatronics*, vol. 20, no. 2, pp. 326–334, 2010.

[31] J. Yu, M. Wang, Z. Su, M. Tan, and J. Zhang, "Dynamic modeling of a CPG-governed multijoint robotic fish," *Adv. Robot.*, vol. 27, no. 4, pp. 275–285, 2013,

[32] J. Yu, M. Wang, M. Tan, and J. Zhang, "Three-dimensional swimming," *IEEE Robot. Autom. Mag.*, vol. 18, no. 4, pp. 47–58, 2011,

[33] K. Matsuoka, "Sustained oscillations generated by mutually inhibiting neurons with adaptation," *Biol. Cybern.*, vol. 52, no. 6, pp. 367–376, 1985,

[34] G. J. Li, L. Zhu, and X. Y. Lu, "Numerical studies on locomotion perfromance of fish-like tail fins," *J. Hydrodyn.*, vol. 24, no. 4, pp. 488–495, 2012,

[35] M. Gazzola, M. Argentina, and L. Mahadevan, "Gait and speed selection in slender inertial swimmers," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 13, pp. 3874–3879, 2015,

[36] Y. Nakamura, T. Mori, M. aki Sato, and S. Ishii, "Reinforcement learning for a biped robot based on a CPG-actor-critic method," *Neural Networks*, vol. 20, no. 6, pp. 723–735, 2007,

[37] M. Ishige, T. Umedachi, T. Taniguchi, and Y. Kawahara, "Exploring Behaviors of Caterpillar-Like Soft Robots with a Central Pattern Generator-Based Controller and Reinforcement Learning," *Soft Robot.*, vol. 6, no. 5, pp. 579–594, 2019,

[38] F. Sehnke, C. Osendorfer, T. Rückstieß, A. Graves, J. Peters, and J. Schmidhuber, "Parameter-exploring policy gradients," *Neural Networks*, vol. 23, no. 4, pp. 551–559, 2010.

[39] Y. Shan, Y. Bayiz, and B. Cheng, "Efficient Thrust Generation in Robotic Fish Caudal Fins using Policy Search," *IET Cyber-Systems Robot.*, vol. 1, no. 1, pp. 38–44, 2019.

[40] J. Peters and S. Schaal, "Policy gradient methods for robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2006, pp. 2219–2225.

[41] T. Zhao, H. Hachiya, G. Niu, and M. Sugiyama, "Analysis and improvement of policy gradient estimation," *Neural Networks*, vol. 26, pp. 118–129, 2012,

[42] C. S. Wardle, J. J. Videler, and J. D. Altringham, "Tuning in to fish swimming waves: body form, swimming mode and muscle function," *J. Exp. Biol.*, vol. 198, no. 8, pp. 1629–1636, 1995.

[43] M. J. Lighthill, "Note on the swimming of slender fish," *J. Fluid Mech.*, vol. 9, no. 2, pp. 305–317, 1960.

[44] N. A. Battista, "Diving into a Simple Anguilliform Swimmer's Sensitivity," *Integr. Comp. Biol.*, vol. 60, no. 5, pp. 1236–1250, 2020,

[45] M. J. Lighthill, "Hydromechanics of aquatic animal propulsion," *Annu. Rev. Fluid Mech.*, vol. 1, no. 1, pp. 413–446, 1969.