Crawling Posture Learning in Humanoid Robots using a Natural-Actor-Critic CPG Architecture

Cai Li, Robert Lowe and Tom Ziemke
Interaction Lab
University of Skovde
Skovde, Sweden
Email: gauss.lee, robert.lowe, tom.ziemke@his.se

Abstract

In this article, a four-cell CPG network, exploiting sensory feedback, is proposed in order to emulate infant crawling gaits when utilized on the NAO robot. Based on the crawling model, the positive episodic natural-actor-critic architecture is applied to learn a proper posture of crawling on a simulated NAO. By transferring the learned results to the physical NAO, the transferability from simulation to physical world is discussed. Finally, a discussion pertaining to locomotion learning based on dynamic system theory is given in the conclusion.

Introduction

Crawling, as a type of quadrupedal locomotion, has been investigated on a number of humanoid robot platforms (Degallier et al., 2008) (Li et al., 2011). Compared to bipedal walking, crawling may be more versatile in terms of posture. For example, human beings can actually do standard crawling (knees and hands), bear crawling (feet and hands), crab crawling (upside down bear crawling), leopard crawling (a military specific crawling with elbows) (Wikipedia, 2013). Moreover, each of the crawling types differs regarding posture and needs specific training to perform. In this article, the work presented will focus on modelling standard crawling and training the NAO robot to learn the optimal posture.

From a bio-inspired perspective, central pattern generators (CPGs) are widely used in modelling locomotion on different robotic platforms (Ji, 2008) (Harischandra et al., 2011) (Zhao et al., 2012), including humanoids (Degallier et al., 2008) (Endo et al., 2008). However, none of the above-cited work involves the posture optimization/learning for crawling and yet the posture determines the type of crawling. As far as CPG architectures are concerned, posture adjustment also plays an important role in the adaptability of CPGs (Grillner et al., 2008) (Orlovsky et al., 1999). Grillner et al. (Grillner et al., 2005) not only posit that sensory feedback and postural control interactively connect the perceived environmental change to human neural structures via CPGs but also point out brainstem and basal ganglia are the two main brain-related parts implicated in the adaptive solutions for locomotion. The latter theoretically implies the role of reinforcement learning (RL) as one possible implementable solution for presenting locomotion capabilities on humanoids.

RL provides an agent the capabilities of learning based on the interaction of the body, the environment and the neural structure. RL is one type of affective-modulation mechanism based on searching solutions for maximizing the reward-related value function (Sutton and Barto, 1998). Grillner et al.’s (Grillner et al., 2008) brain-based perspective on locomotion implies the role of RL in the adaptation of CPGs. Biologically, some scientists assume it accounts for the functions of basal ganglia in human brains (Wiering and van Otterlo, 2012). This neural-anatomic link hints that RL might also be implicated in locomotion and specifically, learning to crawl. Recently, the emerging new methods of continuous-space learning links machine learning closely to RL, such as continuous action space learning automaton (Cacla) and natural actor-critic (NAC) (van Hasselt and Wiering, 2007) (Peters and Schaal, 2008). Both of these methods are used in different motor learning tasks on robots to update in a relatively high-dimensional parameter space (Kober et al., 2012) (Farkas et al., 2012). In our work, because of relatively high learning efficiency of NAC, it is used to figure out the optimal posture of crawling under the guidance of a specific reward function. RL, as an approach independent from traditional supervised and unsupervised learning, is a method which can seamlessly integrate scaffolding-related (supervised) and self-learning factors (unsupervised) into one process.

In this article, a posture learning architecture on a crawling humanoid is presented. In section 2, the main principles/theories pertaining to emulating infant crawling on a humanoid robot are introduced in detail, including the theories of CPGs and RL. In section 3, with the results from the simulation, the learned models are transferred to the physical robot for verification. The statistical results are given and analyzed. In section 4, a conclusion related to RL and learning locomotion is drawn.
Methods and Theories

CPGs and Crawling

Since crawling is a periodic motion, using CPGs to interpret and model it falls within the scope of the biological explanation of locomotion(Grillner et al., 2008)(Scott L Hooper, 2001). Via transferring this knowledge to robotics, it offers new methodologies for understanding morphology constrained locomotion principles on different robots(Ijspeert, 2008), including crawling robots(Harischandra et al., 2011)(Degallier et al., 2008). The approaches for modelling crawling can be classified into two categories: the engineering and the bio-inspired. Specifying the pre-defined trajectories of end-effectors, Aoi et al(Aoi and Tsuchiya, 2005) propose an engineering-based crawling model implemented on a bipedal robot. A key drawback of this approach is that this crawling robot might need recalculation and even remodeling when the environment changes, which limits motoric adaptation capabilities. On the basis of this, a lot of bio-inspired roboticians posit other solutions based on CPGs for avoiding the limitations of engineering methods. Nakamura et al, Righetti et al and Li et al(Nakamura et al, 2007)(Righetti and Ijspeert, 2008)(Li et al., 2013) together applied CPGs to locomotion by regarding sensory feedback as the input to reshape the output of neural controllers, endowing their architectures with self-adaptation capabilities. In this article, the CPG architecture based on sensory feedback reshaping is coined as least sensory feedback CPG model. The definition is as follows:

Least Sensory Feedback CPG Model: The adaptation and feasibility of CPG models rely on a limited/necessary number of sensors so that the output of the model can be interactively reshaped by perceived contextual change. The CPG architecture with above-mentioned characteristics is defined as the least sensory feedback CPG.

“least” here emphasizes necessary/minimal. For example, based on dynamic systems theory, Righetti et al propose the four-cell CPG network with necessary fast/slow transition feedback to implement standard crawling on the iCub robot(Degallier et al., 2008). Without fast/slow transition, the iCub cannot crawl in the real world. Inspired by Righetti et al’s model, we adopt the fast/slow transition into our experiments as a necessary self-modulation mechanism of CPGs.

the mathematical model of CPGs  Inspired by work on iCub crawling(Degallier et al., 2008) and adapted from our previous work(Li et al., 2011), the Hopf oscillators are adapted into our work:

\[
\dot{x}_i = a(m - x_i^2 + y_i^2)y - \omega_i x_i
\]

Methods and Theories

CPGs and Crawling

Since crawling is a periodic motion, using CPGs to interpret and model it falls within the scope of the biological explanation of locomotion(Grillner et al., 2008)(Scott L Hooper, 2001). Via transferring this knowledge to robotics, it offers new methodologies for understanding morphology constrained locomotion principles on different robots(Ijspeert, 2008), including crawling robots(Harischandra et al., 2011)(Degallier et al., 2008). The approaches for modelling crawling can be classified into two categories: the engineering and the bio-inspired. Specifying the pre-defined trajectories of end-effectors, Aoi et al(Aoi and Tsuchiya, 2005) propose an engineering-based crawling model implemented on a bipedal robot. A key drawback of this approach is that this crawling robot might need recalculation and even remodeling when the environment changes, which limits motoric adaptation capabilities. On the basis of this, a lot of bio-inspired roboticians posit other solutions based on CPGs for avoiding the limitations of engineering methods. Nakamura et al, Righetti et al and Li et al(Nakamura et al, 2007)(Righetti and Ijspeert, 2008)(Li et al., 2013) together applied CPGs to locomotion by regarding sensory feedback as the input to reshape the output of neural controllers, endowing their architectures with self-adaptation capabilities. In this article, the CPG architecture based on sensory feedback reshaping is coined as least sensory feedback CPG model. The definition is as follows:

Least Sensory Feedback CPG Model: The adaptation and feasibility of CPG models reply on a limited/necessary number of sensors so that the output of the model can be interactively reshaped by perceived contextual change. The CPG architecture with above-mentioned characteristics is defined as the least sensory feedback CPG.

“least” here emphasizes necessary/minimal. For example, based on dynamic systems theory, Righetti et al propose the four-cell CPG network with necessary fast/slow transition feedback to implement standard crawling on the iCub robot(Degallier et al., 2008). Without fast/slow transition, the iCub cannot crawl in the real world. Inspired by Righetti et al’s model, we adopt the fast/slow transition into our experiments as a necessary self-modulation mechanism of CPGs.

the mathematical model of CPGs  Inspired by work on iCub crawling(Degallier et al., 2008) and adapted from our previous work(Li et al., 2011), the Hopf oscillators are adapted into our work:

\[
\dot{x}_i = a(m - x_i^2 + y_i^2)y - \omega_i x_i
\]
based on a triggered delay in the oscillator, is to smoothen the transition from stance phase to swing phase. These two characteristic types of feedback are activated by pressure sensors in the crawling iCub but maintained as invariant features in our work as the NAO robot does not have usable pressure sensors. They are regarded as types of proprioceptive sensory feedback related to the position of the joint. Then the investigation of crawling in NAO is transformed to the problem of how to find a proper posture that allows the robot to crawl with a crawling-featured "limit cycle".

**Standard crawling and its key variables** Crawling, as the first milestone motion ability in human infants(Clearfield, 2004), serves as a cornerstone gateway to learn the body-environment interaction(Kail and Cavanaugh, 2012). Most expert crawlers during infancy move on knees and hands according to Righetti et al’s investigation(Righetti et al., 2008). Hereby, in order to emulate standard crawling on a robot, the knowledge of how infants learn to crawl might be applicable. Clearfield et al(Clearfield, 2004) emphasize the salient role of spatial-temporal memory, especially the "distance", in the process of infant crawling. Adolph et al(Adolph et al., 2012) found that the parental scaffolding by holding up crawling infants offers positive "safety" assurance on the posture. Therefore, it might be necessary to immerse two factors (distance and posture) into the learning mechanism. The posture can be embodied as the angle affecting the head direction (Figure 2). In our work, it is controlled by a Gaussian distribution between 30 degree and 50 degree.

The number of joints for a humanoid pertaining to the whole-body posture can be very high dimensional. To avoid high dimensionality, firstly, the robot begins in a crawling posture, including shoulders (pitch and roll), hips (pitch and roll) and elbows (roll). The knee and ankle joints are neglected as they do not move so much for crawling(Righetti et al., 2008).

**Natural Actor Critic in Posture Learning**

**Episodic Natural Actor Critic (eNAC)** eNAC is well-known for its learning efficiency on searching optima in a continuous parameterized space. As the posture adjustment requires a continuous adaptation, eNAC is selected to optimize the crawling posture. Compared to Cacla, another efficient continuous-space RL, eNAC might suffer in possible failures by updating a parameterized model into an unknown action space(Wiering and van Otterlo, 2012). So a positive eNAC is applied in our work (for details see part 2) to highlight the potential failures.

NAC was proposed by Kakade et al(Kakade, 2001) and further developed and used in motor learning by Peters et al(Peters and Schaal, 2008). It transforms the traditional RL problem on solving the Bellman equation to an explorative process of linear regression. As a policy gradient approach, the principles of NAC are elucidated as follows:

Assume the stationary policy is \( \pi^\theta(x,u) \) which can determine action space \( u \) based on state space \( x \) with a static distribution \( d^\theta(x) \). The immediate reward is \( r(x,u) \). Then the expected reward \( J(\theta) \) can be written as:

\[
J(\theta) = \int_x d^\theta(x) \int_u \pi^\theta(u|x) r(x,u) dx du
\]

\[
\theta_{n+1} = \theta_n + \alpha \nabla_{\theta} J|_{\theta=\theta_n}
\]

where the policy \( \pi^\theta(x,u) \) is derivable at the policy parameters \( \theta \), namely \( \nabla_{\theta} \pi^\theta \) exists. For maximizing expected reward \( J(\theta) \) with respect to \( \theta \), the policy gradient will find the steepest increase direction \( \nabla_{\theta} J = J(\theta + \Delta \theta) - J(\theta) \) to update searching policy \( \pi^\theta(x,u) \) until it converges. \( \alpha \) represents the n-th step of the update and \( \alpha \) is the learning rate (equal to 0.01). By and large, Equation (6) and (7) plot the rudimentary rule of thumb for policy gradient approaches. Transformed to natural policy gradient, Equation (6) and (7) can be revised to (8) and (9):

\[
\nabla_{\theta} J(\theta) = \int_x d^\theta(x) \int_u \pi^\theta(u|x) \nabla_{\theta} \log(\pi^\theta(u|x))
\]

\[
\nabla_{\theta} \log^T(\pi^\theta(u|x)) \omega dx du = F_{\theta} \omega
\]

\[
\theta_{n+1} = \theta_n + \alpha F^{-1}_{\theta} \nabla_{\theta} J|_{\theta=\theta_n}
\]

\[
F_{\theta} = \int_x \pi^\theta \nabla_{\theta} \log \pi^\theta \nabla_{\theta} \log_\theta d\theta
\]

where \( F \) is the Fisher Matrix (FM). Multiplied by FM, normal policy gradient is changed to the steepest one (Here all the \( x, u \) are neglected for simplification reason). \( \omega \) is a weight vector of the linear approximation and \( \nabla_{\theta} \log^T(\pi^\theta(u|x)) \) is the group of basis functions. Then conclusively, by replacing \( \nabla_{\theta} J(\theta) \) in (9) with (8), the natural PG becomes:

\[
\theta_{n+1} = \theta_n + \alpha \omega
\]

The RL problem changes from searching the steepest policy gradient to a normal regression problem with basis functions. Since the state-action function \( Q^\theta(x,u) = b(x) + \nabla \log(\pi^\theta(u|x)) \omega \) and \( Q^\theta(x,u) = r(x,u) + \lambda \int_{x'} p(x'|x,u) V(x') dx' \) (where \( \lambda \) is the discounting factor, \( x' \) is the next state, \( p(x'|x,u) \) is the probability of state transition), assume the value function is \( V(x) = b(x) \) and can be approximated by \( \psi^T(x)v \) (where \( v \) is the weight vector and \( \psi \) is the vector of basis function related to the value function).(Baird, 1993). Therefore, the approximation can be re-written:

\[
\nabla \log^T(\pi^\theta(u_i|x_i)) \omega + \psi^T(x_i)v = r(x_i,u_i) + \lambda \psi^T(x_{i+1})v + e(x_i,x_{i+1},u_i)
\]
if we sum up left and right side of equation (11), we get a straightforward regression problem for H episodes with s time-length:

$$(\sum_{t=1}^{s} \alpha_t \nabla log^T(\pi^\theta(\mathbf{u}_t | \mathbf{x}_t)))_{1:H} \mathbf{w} + J = \sum_{t=1}^{s} \alpha_t r(x_t, \mathbf{u}_t)_{1:H}$$

where J is the value-function related term considered as a constant baseline and $\alpha_t$ is the average discounting factor. By means of the least square learning rule, the natural PG $\mathbf{w}$ can be obtained for H episodes:

$$\begin{bmatrix} \mathbf{w} \\ J \end{bmatrix} = (\phi \phi^T)^{-1} \phi R.$$

$$\phi = \left[ \sum_{t=1}^{s} \alpha_t \nabla log^T(\pi^\theta(\mathbf{u}_t | \mathbf{x}_t), 1) \right]_{1:H}^T$$

$$R = \left[ \sum_{t=1}^{T} \alpha_t r(x_t, \mathbf{u}_t) \right]_{1:H}^T$$

where $1 : H$ represent H times samplings within one trial (refer to details in the Algorithm).

**Implementation of the learning algorithm** Based on the above-mentioned proof (for details, please refer to Peters, 2007), the state space and action space are defined as $P \sim [P_{\text{shoulder}}, P_{\text{hip}}, P_{\text{elbow}}]$ and $U \sim [U_{\text{shoulder}}, U_{\text{hip}}, U_{\text{elbow}}]$. So $P = P_0 + U$, where $P_0$ is the initial posture vector. With the parametrized policy $\pi^\theta(U|P)$, a posture can be learned via the optimization of the gaussian policy:

$$\pi^\theta(U|P) = N(U, \bar{U}, \sigma) = \frac{2\pi}{\sigma} \exp \left( \frac{(U - \bar{U})(U - \bar{U})^T}{\sigma^2} \right)$$

where $U$ is the output vector of the policy and $\bar{U}$ is the input vector equal to the updated $U$ from last trial and $\bar{U} \sim [\theta_{\text{shoulder}}, \theta_{\text{hip}}, \theta_{\text{elbow}}]$. $\sigma$ is the exploration rate which determines the variance of U from U. In our work, $\sigma = 0.03$ as if $\sigma > 0.05$, the posture updating is unstable and if $\sigma < 0.01$, the posture adjustment is too insensitive and time-consuming.

The schema of eNAC algorithm is shown below:

**eNAC Algorithm:**

Repeat M trials each of which includes 10 rollouts (H=10), use the policy $U_{1:H} \sim \pi^\theta(U|P)$ to generate H groups of actions, each action is taken for time $t=1,2...s$.

Calculate: for each rollout, the episodic return $r_i =$

$$\sum_{t=1}^{s} \alpha_t r(x_t, \mathbf{u}_t)$$

the eligibility $\psi_t = \nabla log^T(\pi^\theta(\mathbf{u}_t | \mathbf{x}_t), 1)$

then the gradient is:

$$\begin{bmatrix} \mathbf{w} \\ J \end{bmatrix} = (\phi \phi^T)^{-1} \phi R.$$

where $R = [r_1, r_2, ..., r_H]^T$ and $\phi = [\psi_1, \psi_2, ..., \psi_H]^T$

Updating for each trial:

if $\delta > 0$, with $\delta = R_{\text{avg}} - V_n$ where $R_{\text{avg}}$ is the average of R and $V_n$ is the episodic value function of last updating:

$$V_{n+1} = V_n + 0.1 \times \delta$$

and $\theta_{n+1} = \theta_n + \alpha \mathbf{w}$, otherwise no updating.

Until the convergence condition is satisfied: $\delta < 0$ all the time or $\delta < 10^{-4}$.

It is noteworthy that, inspired by the Cacla architecture (van Hasselt and Wiering, 2007), the “positive updating” is used to avoid the inappropriate updating in the parameterized action space. Since the function approximation cannot accurately converge to the real Q function ($\mathbf{w}$ cannot be zero), the convergence condition is necessary to determine the termination of each learning process.

**Experimental Setup**

As mentioned in Section 1, the crawling distance and spineline angle are the two important variables for evaluating the crawling posture (Figure 2). Accordingly, in RL, the reward function is composed of two terms ($r_{\text{distance}}$ and $r_{\text{angle}}$) related to the two factors that represent the two above-mentioned variables:

$$r_{\text{reward}} = r_{\text{distance}} + r_{\text{angle}}$$

$$r_{\text{distance}} = \exp\left(\frac{D}{2}\right) - 1$$

$$r_{\text{angle}} = \exp(e) - 1$$

with $e = N(x_0 = 1.05, \sigma = 0.1)$

where $D$ is the distance the robot crawls every episode. $e$ is a Gaussian distribution with the center 1.05 (approximately 45 degree) and variance 0.1. As a matter of fact, distance is a very important measure of the improvement of crawling and spineline angle is used to control and stabilize the body-height alteration.

In the experiments described below, the crawling CPGs are firstly implemented on a simulated NAO robot in Webots which is a popular commercialized robot simulator (Michel, 2004). Then the statistically-learned models are transferred to the physical robot for testing. The advantage of using the simulated robot is not only to avoid unexpected damage upon the physical robot but also to simplify the measurement of distance by using the special supervisor functions of Webots.

1185 ECAL 2013
Figure 2: The standard crawling posture on knees and hands. The distance and spineline angle indicate the quality of crawling. The spineline angle is limited in a scope (30 degree~60 degree) by a Gaussian function.

**Experiment 1: Utilizing proprioceptive sensory feedback**

This experiment is to verify the importance of proprioceptive sensory feedback (PSF) characterizing the crawling CPG. The snapshots of our crawling video (Figure 3) shows the utility of PSF. In this experiment, the NAO robot crawls with a manually-tuned posture. From the video(Li, 2013b), it is clearly observed that the NAO robot can crawl much further/faster with PSF. The multiple runs do not qualitatively effect the performance of crawling with possible simulated noise effects.

**Experiment 2: Learning the posture of standard crawling**

With the crawling CPG and its proprioceptive sensory feedback, eNAC is applied on the simulated NAO robot for learning the optimal posture. Each episode of eNAC lasts 25s and the posture is continuously optimized by maximizing the crawling distance (in the same time duration) within the scope of the spineline angle. Each run of the experiment takes approximately 7 hours to converge (on a single MacBook pro). In total, 10 results are obtained statistically by running the experiment with variable values reset at the beginning of each run 10 times. Figure 4 shows the snapshots of optimized crawling and its comparison to the original one. Based on the video(Li, 2013c) and the snapshots, we can summarize the observable difference of two crawling gaits before/after learning. Firstly, before learning, the robot struggles to crawl forward with an obvious slippage on hands but after learning, the slippage disappears. Secondly, the crawling distance has been well optimized after learning. Subsequently, a detailed understanding of how the posture is adjusted by the eNAC is explained in the next section.
Results and Analysis

Results analysis

Since every run of each experiment starts independently from the same initial posture, it is intriguing to see how the statistical learning process tunes the six joints together and also if the 10 results are consistent. The typical characteristic of eNAC is that it can find the correct updating directions in the parameter space with a relatively high learning efficiency. Even though it might suffer in leading the updating into uncertain/invalid directions (Wiering and van Otterlo, 2012), the positive eNAC used in this article could possibly avoid this drawback. Figure 7.A~B show the increase of reward during the 10-run experiment. All 10 runs show the reward levels are boosted based on postural and distance rewards. Moreover, in almost every run of the experiment, the reward is effectively augmented during the first 20 episodes. This implies that positive eNAC could seek out valid updating pathways in a short time. From the other point of view, the details of joint-value updating are highlighted in the Figure 7.C~H. Generally, even though different runs converge to distinct results (marked by black circles) within disparate updating episodes, the updating curves of each joint in every run share qualitatively similar update profiles and converge towards same directions. Therefore, above-mentioned conclusive summary insinuates the consistency over independent runs of learning processes. As a statistical learning approach, all the results obtained in the 10-run experiment are not the same. However, the converged posture adjustment values are located in a close scope. For example, $\hat{U} \sim [\theta_{\text{HipYawPitch}}, \theta_{\text{HipRoll}}, \theta_{\text{HipPitch}}, \theta_{\text{ShoulderPitch}}, \theta_{\text{ShoulderRoll}}, \theta_{\text{ElbowRoll}}]$ converges to the values approximately around 0.05 rad, 0.1 rad, -0.15 rad, -0.1 rad, -0.15 rad, 0.1 rad respectively (observed from Figure 7.C~H).

Figure 5 displays the improvement of crawling gait in two aspects. Compared to the initial state, the crawling distance and spineline angle are both improved. In this article, the standard crawling posture is maintained by $r_{\text{angle}}$ which defines the scope of spineline angle. It somewhat stabilizes the crawling posture adjustment. In the experiment, it is observed that if the angular feedback mechanism is

omitted, the robot produces a crawling behaviour (flat scorpion-like motion) that does not conform to any typical infant crawling gaits (Figure 6). Although the scorpion-like crawling is an optimized solution in terms of only the distance, it is still considered to be an improper result as non-standard crawling posture. The crawling distance (in the same time duration) is a measure indicating the embodied improvement of crawling gait. The two different rewards are two reinforced drivers to make a proper crawling gait emerge.

Transferrability to the physical robot

The 10 results obtained via eNAC are transferred to test on the physical robot. The transferrability from the simulated robot to the physical robot is always a significant problem. The main reasons causing the failure of transferring are: firstly, the inappropriate physics engine might lead to distinct physical interaction in the simulator compared to the real world, including the collision detection which is very important for locomotion modelling. Webots uses a widely-used physics engine, ODE (Open Dynamics Engine (Michel, 2004)). It has attained reliable performance in robotic simulation, especially on locomotion-related tasks (Harischandra et al., 2011) (Degallier et al., 2008). Secondly, the difference in timing. The machine time might be different from the real time. In our work, the physical robot cannot crawl with the straightforward transferring from the simulated results. However, the crawling gait is successfully transferred by doubling the CPG frequency. It seems the numerical integration used to calculate CPG has a strong attachment to the speed of the machine, causing a difference in the CPG timing on different hardware. Therefore, in order to transfer the work in the simulator to the physical robot, the timing relation between the simulation computer and robot’s hardware has to be identified.

Figure 8 shows the gait performance on the physical robot with one crawling step (for the video, refer to (Li, 2013a)). Authors tested 10 results of the learning. 7 out of 10 can successfully perform a smooth crawling gait. The other 3 failed as the low value of $\theta_{\text{ShoulderPitch}}$ cannot allow the arm to lift above the ground completely when the two arms are alternating. The robot cannot crawl “smoothly” even though it can crawl nicely in the simulator. This might be caused by the difference between the simulated world based on the
Figure 7: Figure 7.A~B show the reward increase of 10-run experiments. The "updating episodes" is the episode which satisfies the updating condition of eNAC. Figure 7.C~H depict the change of joint values during the learning process and it covers the results of 10 runs for each joint. Those joints include HipYawPitch, HipRoll, HipPitch, ShoulderPitch, ShoulderRoll and ElbowRoll, the six joints affecting the quality of crawling gait. The black circles mark the converged points of each joint.

Figure 8: The implementation on the physical robot. This figure shows the video snapshot of one-crawling-step NAO robot on a wooden flat table (One crawling step means one time alternation of the supporting leg and arm.).
Conclusion

In the previous two sections, we verified the necessity of proprioceptive sensory feedback and optimized postures for the crawling gait and tested the gait on the physical robot successfully. The results of experiments highlight the importance of sensory feedback and postural control for the CPG architecture. With a complete least-sensory-feedback CPG, the sensory feedback is designed to reshape the CPG output and the posture control is involved in shifting the oscillation centers of CPGs. Grillner et al.’ work also implies these two functions in the CPG architecture(Orlovsky et al., 1999)(Grillner et al., 2008). Using eNAC, it assists the robot to learn/adapt on the ground of the interaction with the environment.

Conclusion

From Grillner’s Inspiration towards Reality

Grillner et al.’s work on the theories of CPGs might be one of the most complete collections for CPG “huggers”. His group not only investigate the biological theories of CPGs spanning from mollusc to man(Orlovsky et al., 1999) but also propose the plausible architecture of a goal-directed locomotion system which links CPGs to brain in an exact way(Grillner et al., 2005)(Grillner et al., 2008). This connection between the brain and CPGs is read as the connection between RL and oscillators in this article. RL fulfills the functions of basal ganglia and MLR (mesencephalic locomotor region) highlighted by Grillner et al(Grillner et al., 2008). The roles of CPGs include sensory-feedback integration and posture control. With the least-sensory-feedback concept, in this article, authors simply interpret roles of CPGs as the output reshaping and postural learning. Even though the crawling posture is successfully learned in our work, several questions still remain unsolved: Firstly, the least-sensory-feedback CPG model might not be the best solution to explaining sensory feedback integration. With predefined sensory feedback, even though the CPG model can acquire the adaptation in a lot of work(Li et al., 2011)(Li et al., 2012)(Li et al., 2013)(Nassour et al., 2013)(Nakamura et al., 2007), it is still functionally limited. If the sensor configuration is altered or the robot lacks necessary sensors (the case in this article), the CPG architecture might lose its adaptation to the changing environment. Therefore, a general solution for integrating sensory feedback may be demanded. Secondly, the energy efficiency of CPGs. Energy efficiency is a highlighted property of using CPGs(Tomoyuki et al., 2009)(Matsuo et al., 2012). Even though CPGs work as trajectory generators in our work, the energy efficiency is actually, at least, not high as the stiffness of joints is maintained as constant 0.9 (0~1.0 for the NAO robot). The link between energy efficiency and CPGs seems not be clear in theories. It might be a factor also related to morphologies. Since the NAO robot has a rigid body, it might be difficult to achieve high energy efficiency.

Rethinking locomotion learning

The focus of locomotion modelling has been gradually expanded into bio-inspired approaches(Ijspeert, 2008). Pfeifer et al emphasize the role of body morphologies as a core value of soft robotics(Pfeifer and Bongard, 2006). The locomotion learning/modelling has been elucidated as an interactive process instead of only being based on engineering approaches. The interaction of neural controllers, the body and the environment is the core of this ideology. From this angle, RL, as an interactive learning process, is able to bridge the three parts. Especially, with the newly proposed RL approached in continuous space, it can serve locomotion learning for more complicated parameterized models. This not only fits within the perspective of the architecture of Grillner et al.’s proposition(Grillner et al., 2008) but also is consistent with the theories of dynamic systems theory(Thelen, 1996).

Drawbacks and Future work

Aside from the two major drawbacks mentioned above, the transferability from simulation to physical world is also an unresolved problem. In order to avoid this problem, the learning process should be able to transfer to the physical robot. In our work, the learning approach relies on the measurement of crawling distance. Webots proffers a convenient solution on any distance measuring with supervisors. In the physical world, the robot should have the ability to measure distance visually so that the learning process can be transferred. In future work, the motor primitive models(Kober et al., 2012) might be used to form a general solution to reshaping the CPG output. With this approach, the least-sensory-feedback model can possibly be adapted into a generic CPG model.

Acknowledgment

The authors would like to thank the European Project RobotDoc (http://www.robotdoc.org) for funding and providing help towards our work.

References


Li, C. (2013a). crawling before learning and after learning: https://www.youtube.com/watch?v=5mygX30x96U.


