# Behaviors Coordination and Learning on Autonomous Navigation of Physical Robot

**Handy Wicaksono*[1], Handry Khoswanto[1], Son Kuswadi[2]**
[1]Electrical Engineering Department, Petra Christian University
[2]Mechatronics Department, Electronics Engineering Polytechnic Institute of Surabaya
e-mail: handy@petra.ac.id[1], handry@petra.ac.id[2], sonk@eepis-its.edu[3]

### Abstrak

  *Koordinasi perilaku adalah salah satu faktor kunci pada robot berbasis perilaku. Arsitektur subsumption dan motor schema adalah contoh dari metode tersebut. Untuk mempelajari sifat keduanya, eksperimen pada robot fisik perlu dilakukan. Dari hasil eksperimen dapat disimpulkan bahwa metode pertama memberikan respons yang cepat, robust tetapi tidak halus. Sedang metode ke dua memberikan respons yang lebih lambat namun lebih halus, dan cenderung menemukan target lebih cepat. Perilaku yang mampu belajar dapat memperbaiki performa robot dalam menghadapi ketidakpastian. Q learning adalah metode pembelajaran reinforcement yang populer digunakan pada pembelajaran robot karena sederhana, konvergen dan off policy. Variabel laju pembelajaran berpengaruh pada performa robot dalam fase pembelajaran. Algoritma Q learning diterapkan pada subsumption architecture dari suatu robot fisik. Sebagai hasilnya, robot telah berhasil melakukan navigasi otonom meski dengan beberapa keterbatasan akibat peletakan dan karakteristik sensor.*

**Kata kunci:** *behavior based robotics, coordination, reinforcement learning, navigasi otonom*

### Abstract

  *Behaviors coordination is one of keypoints in behavior based robotics. Subsumption architecture and motor schema are example of their methods. In order to study their characteristics, experiments in physical robot are needed to be done. It can be concluded from experiment result that the first method gives quick, robust but non smooth response. Meanwhile the latter gives slower but smoother response and it is tending to reach target faster. Learning behavior improve robot's performance in handling uncertainty. Q learning is popular reinforcement learning method that has been used in robot learning because it is simple, convergent and off policy. The learning rate of Q affects robot's performance in learning phase. Q learning algorithm is implemented in subsumption architecture of physical robot. As the result, robot succeeds to do autonomous navigation task although it has some limitations in relation with sensor placement and characteristic.*

**Keywords**: *behavior based robotics, coordination, reinforcement learning, autonomous navigation*

## 1. Introduction

  Behavior based architecture is a key concept in creating fast and reliable robot. It replaces deliberative architecture that used in Shakey robot [1]. Behavior based robot doesn't need world model to finish its task. The environment is the only model needed. Another advantage is all behaviors run in parallel, simultaneous, and asynchronous way [2]. In this architecture, robot must have behavior coordinator to coordinate robot's behaviors. First approach suggested by Brooks [2] is Subsumption Architecture that can be classified as competitive method. In this method, there is only one behavior that can be applied in robot at one time. It is very simple and it gives the fast performance result, but it has disavantage of non-smooth response and inaccurate. To overcome competitive method weakness, Arkin [3],[4] suggests Motor Schema that can be classified as cooperative method. In this method, there can be more than one behavior that applied in robot at one time so every behavior has contribution in robot's action. This method results in smoother response and more accurate, but it is more complicated. The complete list of behavior coordination methods can be found in [5].

  In order to anticipate many uncertain things, robot should have learning mechanism. In supervised learning, robot will need a master to teach it while unsupervised learning mechanism will make robot learn by itself. Reinforcement learning (RL) is an example of this

method, so robot can learn online by accepting reward from its environment [6]. There are many RL applications on robotics, including: free gait generations for six legged robot [7] and robot grasping [8]. There are many methods to solve RL problem. One of most popular method is Temporal Difference Algorithm, especially Q Learning algorithm [9]. Q Learning advantages are its off-policy characteristic and simple algorithm. It is also convergent in optimal policy. But it can only be used in discrete state/action. If Q table is large enough, algorithm will spend too much time in learning process [10].

In order to study characteristics of behaviors coordination methods and behavior learning above, some researchers has done simulations by using robotic simulator software [11], [12]. Simulation is needed because learning algorithm usually takes more memory space on robot's controller and it also adds program complexity. However, experiments with physical robot are still needed to be done, because there are big differences between ideal environment and real world. Robot will accomplish autonomous navigation task by developing adaptive behaviors. Because of limited resources (e.g. sensors), this robot does not have capabilities to build and maintain environment's map. Nevertheless it still ables to finish the certain task [13]. This paper will describe about behavior coordination and learning implementation on physical robot that can navigate autonomously.

## 2. The Proposed Method
### 2.1. Behaviors Coordination
In behavior based robotics approach, methods of behaviors coordination are significant. The designer needs to know how robot coordinate its behaviors and take the action in the real world. There are two approaches: competitive and cooperative. In competitive method, at one time, there is only one behavior that applied in robot.

The first suggestion on this type is Subsumption Architecture that suggested by Brooks [2]. This method divides behaviors to many levels, where the higher level behavior have higher priorities too. So it can subsume the lower level ones. The layered control system is shown on Figure1.
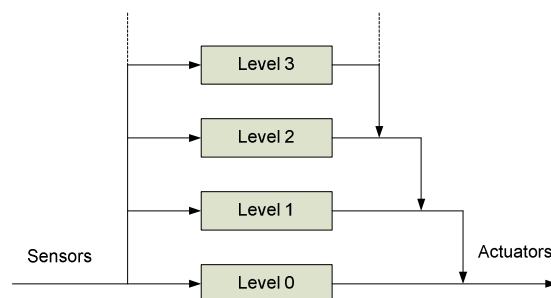


Figure 1. Layered control system [2]

The cooperative method have different approaches. In this method, at one time, there can be more than one behavior that applied in robot, so every behavior has contribution in robot's action. Arkin [3] suggest the motor schema method, which every object will be described as vector that has magnitude and direction. The result behavior is mixing between each behavior. The motor scheme for this method appears on Figure 2. Some experiments will be done to compare the behavior coordination methods implementation on autonomous navigation task of physical robot.

### 2.2 Learning Behavior
Robot using proper configuration of behaviors coordination method will accomplish task given by human well. However, in some unpredictable conditions by human designer, robot should have intelligence to make its own decision. One of learning method that suitable for robot application is reinforcement learning (RL), a kind of unsupervised learning method which

learns from agent's environment [8]. Agent (such as: robot) will receive delayed reward from its environment. Figure 3 shows reinforcement learning basic scheme.
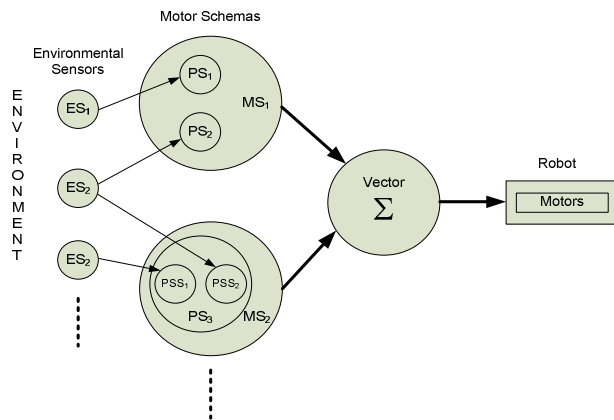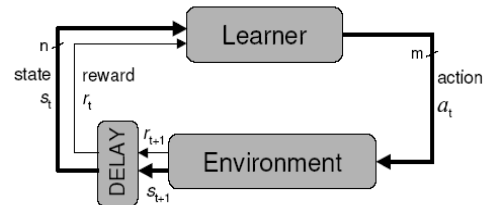


Figure 2. Motor schema method [3]



Figure 3. Reinforcement learning basic scheme [8]

There are some reinforcement learning methods : Sarsa, Actor Critic, Q learning, etc. Q learning is most popular RL method that applied in robotics because it is off policy (others are on policy) and simple [8]. It also has been convergently proofed [9]. Pseudocode of Q learning algorithm is shown below [10].

```
Initialize Q(s,a) arbitralily
Repeat (for each episode) :
        Initialize s
        Repeat (for each step of episode):
        Choose a from s using policy
                derived from Q (e.g., ∈-greedy)
        Take action a, observe r, s'
        Apply
```

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$$

```
        s ← s';
until s is terminal
```

where:
Q(s,a) : component of Q table (state, action)
s: state          s': next state     a: action              a' : next action
r: reward         $\alpha$ : learning rate  $\gamma$ : discount factor

## 2.3 Learning Behavior on Behaviors Coordination

Learning behavior and behaviors coordination are needed by robot to accomplish its task and adapt with unpredictable environment well. Hence, learning behavior needs to be included in of behaviors coordination method. Figure 4 shows proposed method of behaviors coordination which is combine learning behaviors and non learning ones. Some experiments on behaviors coordination that include Q learning behavior will be done. Another contribution of this paper is implementation of this method on physical robot, because usually it is applied in robotics simulation software only [11], [12].

## 3. Research Method
## 3.1 Robot's Behaviors Design

In order to finish autonomous navigation task, robot should have these behaviors: bstacle avoidance, search target, wandering, and stop. Subsumption architecture (as competitive behaviors coordination method) for robot can navigate autonomously shown in Figure 5.
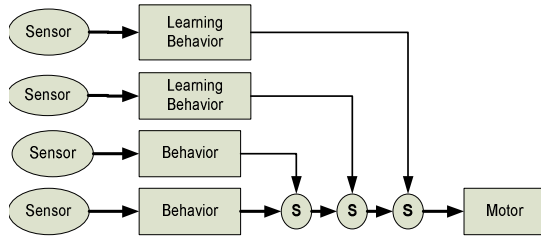
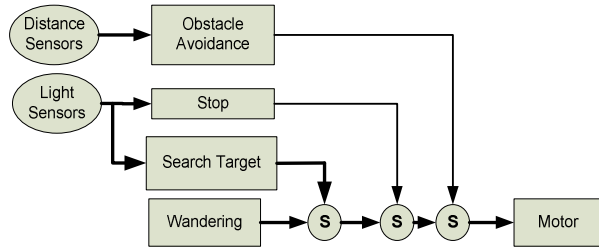Figure 4. Proposed behaviors coordination method



Figure 5. Robot's subsumption architecture for autonomous navigation

Form figure above, it can be seen that robot use distance sensors to detect the obstacle and light sensors to find the target (candle light). Obstacle avoidance become the most important behavior, and wandering become the least one. There is only one behavior that can be used by robot at one time. Pseudocode of this architecture is shown below.

```
IF distance sensors is near the obstacle
    THEN robot avoids the obstacle
ELSE
    {IF light sensors is very near with candle light
        THEN robot is stop
    ELSE
        {IF light sensors is near with candle light
            THEN robot move towards candle light
         ELSE
            Robot is wandering everywhere
        }
    }
```

The example of cooperative behaviors coordination method is motor schema. Its application for robot's autonomous navigation can be shown in Figure 6. The behavior structure is similar with Subsumption Architecture, except the way to mix all robots' behaviors. Here is the pseudocode of architecture above.

```
IF distance sensors is near the obstacle
    THEN compute obstacle avoidance behavior contribution
IF light sensors near with candle light
    THEN compute search target behavior contribution
IF light sensors is near with candle light
    THEN compute stop behavior contribution
Compute wandering wandering behavior contribution

Compute all behaviors contribution and translate it to motor's speed and
direction
```

Design of behaviors coordination method (in example: Subsumption Architecture) that incorporate learning behavior (obstacle avoidance) is shown in Figure 7.
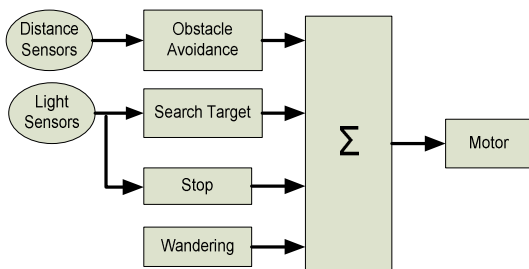


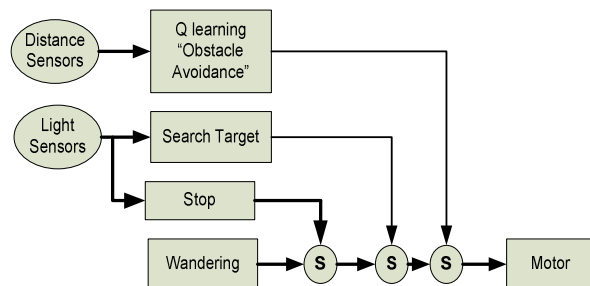Figure 6. Robot's motor schema for autonomous navigation



Figure 7. Q learning behavior on robot's subsumption architecture

Pseudocode of Q learning algorithm is shown on section 2.2, while pseudocode of architecture on Figure 7. is shown below.

```
IF distance sensors is near the obstacle
    THEN robot learns to avoid the obstacle
ELSE
    {IF light sensors is very near with candle light
        THEN robot is stop
    ELSE
        {IF light sensors is near with candle light
            THEN robot learns to move toward candle light
         ELSE
            Robot is wandering everywhere
        }
    }
```

Design of state and reward are important in Q learning algorithm. Here is states value design of robot's obstacle avoidance behavior:
- 0 : if obstacle is far from left and right side
- 1 : if obstacle is near from left side and far from right side
- 2 : if obstacle is far from left side and near from right side
- 3 : if obstacle is near from left and right side

Meanwhile rewards design of the same behavior is:
- 2 : if obstacle is not very near from left and right side
- -1 : if obstacle is very near only from left side or if obstacle is very near only from right side
- -2 : if obstacle very near from left and right side

Other experiments will be done by incorporating search target as Q learning behavior.
State design of this behavior is the same with obstacle avoidance learning behavior:
- 0 : if target is far from left and right side
- 1 : if target is near from left side and far from right side
- 2 : if target is far from left side and near from right side
- 3 : if target is near from left and right side

But rewards design is little bit different with the first behavior. Here it is:
- 4 : if target is very near from left and right side
- -1 : if target is very near only from left side or if obstacle is very near only from right side
- -2 : if target is not very near from left and right side

### 3.2 Physical Robot Implementation

Simulation becomes an important aspect of robotic research. In comparison with real robot experiments, simulations are easier to set up, less expensive, faster, more convenient to use, and allow the user to perform experiments without the risk of damaging the robot [14]. However physical robot experiment is still urgently needed. There still many unpredictable aspects of robot that can not be perfectly modeled by robotics simulation software.

In order to realize physical robot, there are many robotics platform nowadays. Students or researchers don't have to build robot from the beginning, but they can use robotic kit that available on the market today. LEGO NXT Robot is famous robotic kit. It consists of NXT Brick as controller, many kind of sensors (ultrasonic sensor, light sensor, touch sensor and sound sensor), and servo motors as actuator. Nowadays it has been used in advance robotic application such as environment mapping [15], multi robot system [16], [17], robot manipulator [18] and robot learning [19].

This paper will describe about implementation of behavior coordination on LEGO NXT Robot. NXC (Not eXatcly C), an open source C-like language, will be used to program the robot as substitute of NXT-G. There are some NXC programming techniques on implementation of robot's Q learning behavior. Q learning algorithm needs 2 dimensional array to build Q table consist of state action. Enhanced NBC/NXC firmware that support multi dimensional array will be used here. It is also important to use float data type on $\alpha$ (learning rate) and $\gamma$ (discount rate), so their value can be varied between 0 and 1.

LEGO NXT robot used in this research will use two ultrasonic sensors (to detect the obstacles), two light sensors (to detect the target) and two servo motors. NXT Brick behaves as "brain" or controller for this robot. The robot is shown in Figure 8. There are some experiments that will be done here: reaching target, robot's movement, and target versus obstacle experiment. Robot's arena contains some obstacles and one candle as the target. It has three

different home positions. The arena is shown on Figure 9. Other arena with simple structure (by using one obstacle and one target only) will also be used in experiments. They are shown on Figure 10.
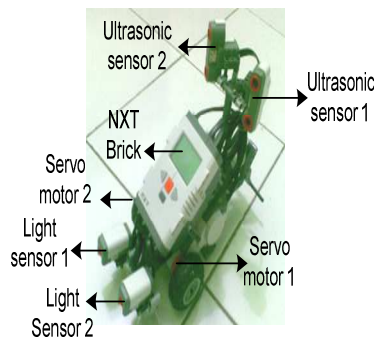
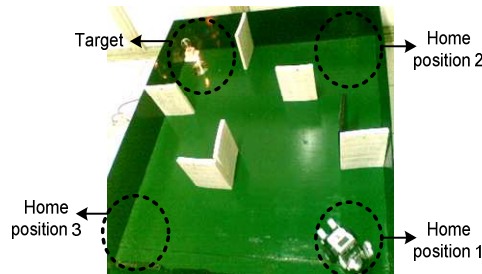Figure 8. LEGO NXT Robot for autonomous navigation task

Figure 9. Complete arena

Figure 10. Simple arena

## 4. Results and Analysis
### 4.1 Reaching the target

This experiment will measure time that needed by robot (with different behavior coordination method) to reach the target. It has been done from three different home positions (see Figure 10.). The result is shown in Table 1. From the table, it can be shown that robot with Motor Schema can reach the target faster than the Subsumption Architecture robot. The reason of this result can be analyzing robot's movement.

### 4.2 Robot's movement in arena

This section will analyze trajectory that made by the robot when it navigate autonomously to find the target. Figure 11 and 12 shows trajectory of Subsumption Architecture and Motor Schema robot from three different positions.

Table 1. Time to reach the target

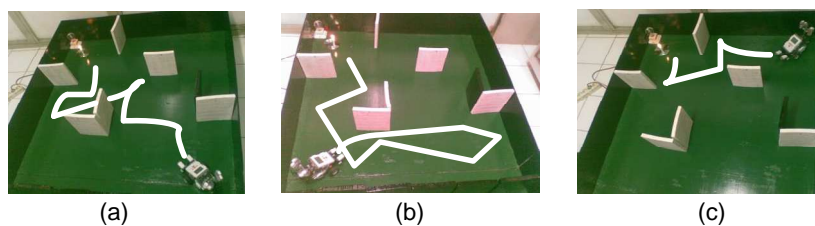| Home Position | Subsumption Architecture (seconds) | Motor Schema (seconds) |
|---|---|---|
| Position 1 | 25 | 20 |
| Position 2 | 47 | 45 |
| Position 3 | 23 | 14 |

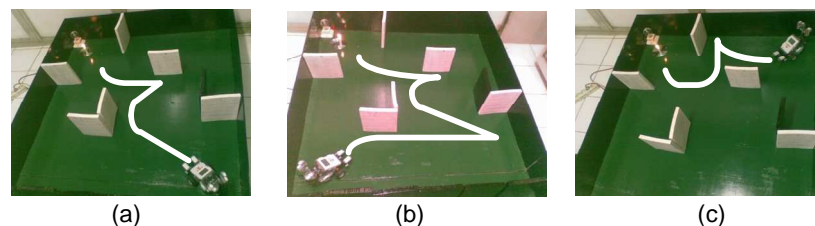Figure 11. Subsumption Architecture robot trajectory from home position 1, 2, and 3

Figure 12. Motor Schema robot trajectory from home position 1, 2, and 3

From those figures above, it can be seen that Subsumption Architecture robot's movement is sharp and not smooth. There are many sharp turn on this robot's trajectory. From the experiments video it appears that this robot is also faster than the other. However on this reaching target behavior, it's not useful much because when the robot move too fast with the sharp movement, the target can be "lost" on the robot's sight. On the other hand, Motor Schema robot's movement is smoother than the preceding one. The sharp turns are not completely lost, but its number is less than before. This robot has slower movement, but it has more accurate detection on target location. That's why the time that needed by this robot to reach the target is faster than the first one.

### 4.3 Target versus obstacle experiment

This experiment will be done to observe robot's characteristics if target and obstacle are located near the robot. Experiment result shown on Figure 13. From figure above it can be seen that Subsumption Architecture give more reactive action by avoiding the obstacle and (at the same time) leaving the target. It is reasonable because obstacle avoidance is the most important behavior of robot. Meanwhile the motor schema robot moves slower, so it can detect the target. It can be happened because robot considers the target location that is near to the robot.

### 4.4 Q learning - obstacle avoidance behavior with fixed learning rate

In this experiment, Q learning is applied in obstacle avoidance behavior only. In order to watch robot's performance, a simple obstacle structure is prepared. Q learning algorithm applied on robot use $\alpha = 0.7$ and $\gamma = 0.7$. It utilizes greedy method for exploration – exploitation policy. Robot's performance on the beginning and the end of trial is shown on Figure 14. and Figure 15.

It can be seen from figures above that robot's learning result can be different between one and other robot. The first robot tends to go to right direction and the second one chooses left direction. Both of them are succeed to avoid the obstacle. This can be happened because Q learning gives intelligence on each robot to decide best action for robot itself.

Robot's goal in Q learning point of view is collect positive rewards as many as possible. Graphic of rewards average every ten iterations and total rewards during the experiment is shown on Figure 16 and Figure 17. From Figure 16, it can be seen that average reward that received by robot is getting better over the time. In the learning phase robot still receive many negative rewards, but after 5 steps it starts to collect positive rewards. Figure 17 shows total (accumulated) rewards collected by robot is getting larger over the time. So it can be concluded that robot can maximize its reward after learning for some time.



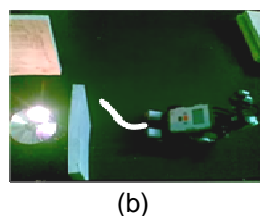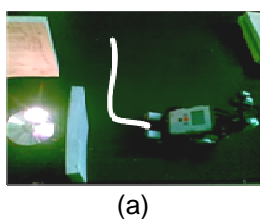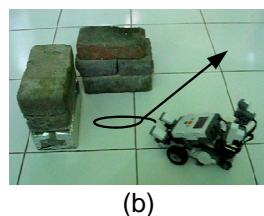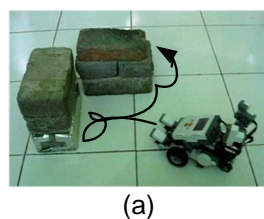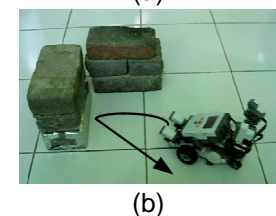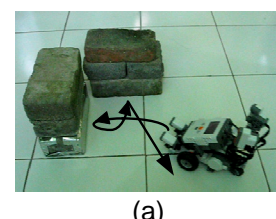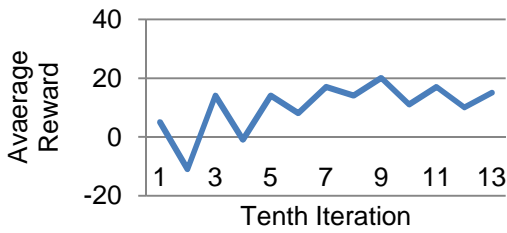| (a) | (a) | (a) |
|-----|-----|-----|
| (b) | (b) | (b) |
| Figure 13. Subsumption Architecture and Motor Schema robot near obstacle and target | Figure 14.  Robot's performance at the beginning and the end of trial 1 | Figure 15.  Robot's performance at the beginning and the end of trial 2 |

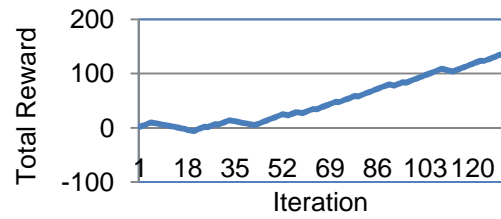Figure 16.  Average reward every tenth iteration



Figure 17.  Total rewards of Q learning obstacle avoidance behavior.

### 4.5 Q learning - obstacle avoidance behavior with varying learning rate

In this experiment, different learning rate (α) will be given to the robot's Q learning algorithm. Its values are: 0.25, 0.5, 0.75 and 1. The result is shown in Figure 18. From Figure 18 (a) and (b), it can be seen that robot with 0.25 (sometimes 0.5) learning rate can not learn to avoid obstacles. But robot with 0.5, 0.75 and 1 learning rate can learn obstacle avoidance task well (it shown on Fig. 18. (c) – (h)). Before robot learns, it will bump to the obstacles sometime because it still doesn't understand that it is forbidden. But after it has learned, it can avoid obstacle (without bumping) successfully.
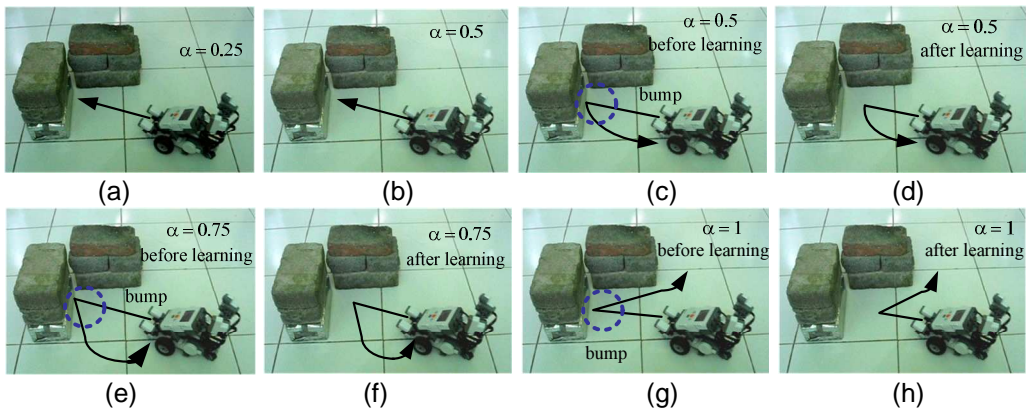


Figure 18.  Robot's movement with different learning rate values

Table 2. Comparison of robot with different learning rate.

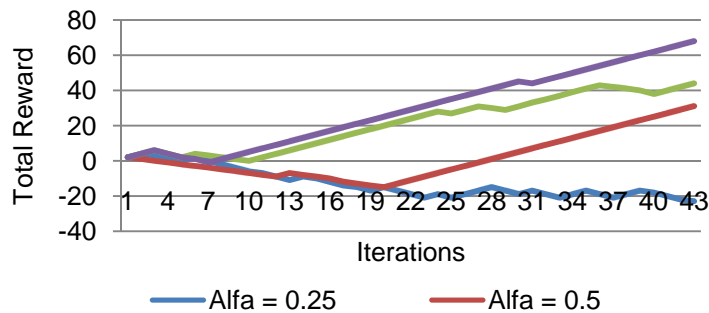| α | Before learning (seconds) | After learning (seconds) |
|---|---|---|
| 0.5 | 15 | 7 |
| 0.75 | 9 | 5 |
| 1 | 7 | 7 |



Figure 19.  Total rewards collected by robot's obstacle avoidance behavior.

      The difference of robot with 0.5, 0.75 and 1 learning rate is time needed to learn and finish obstacle avoidance task. Table 2 is the comparison table of them. From Table 2, it can be seen that the increasing of learning rate is proportional with decreasing time needed by robot to solve the task. In this case, robot with $\alpha = 1$ is the fastest. But in after-learning phase, that robot is not always being the fastest one too.

      Beside the time needed to learn and finish the task, also robot receives different rewards. Amount of rewards collected by robots is shown on Figure 19. From Figure 19, it is shown that robot with bigger learning rate will collect the bigger amount of rewards too. It means that robot will learn the task faster. So it can be concluded that for simple obstacle avoidance task, the best learning rate ($\alpha$) that can be given by robot is 1.

### 4.6 Q learning - search target behavior with fixed learning rate
      In this experiment, Q learning is applied in search target behavior only. Simple arena with one candle as target is prepared to test this behavior. There are two home positions of robot (left and right side of target). The result is shown in Figure 20. From the figure, it can be seen that before learning, robot doesn't know that it should go toward goal (bold line). But after learning, robot will go to where the goal is (dash line). In search target behavior experiment, robot tends to get negative rewards because it doesn't know exactly where the goal is. It is true because RL is kind of trial and error method. So it can be concluded that Q learning application on search target behavior is not suitable for autonomous navigation task. Because of that, robot's performance is not shown by rewards collected by robot, but by amount of iterations robot need to find target (see Figure 21). From Figure 21 it can be shown that after some trials robot is reaching target faster than before.

### 4.7 Q learning - obstacle avoidance behavior on autonomous navigation task
      This Q learning behavior has been used in physical robot that solve autonomous navigation task. Here is the experiment result (see Figure 22.). This robot succeed to avoid the obstacle (after some learning time) and reach the target (by its combination with search target behavior), but it also has some weaknesses. Dashed rectangle in the figure shows some physical problems of light sensor placement in robot and ultrasonic sensor characteristics. Figure 23 describes those physical problems.
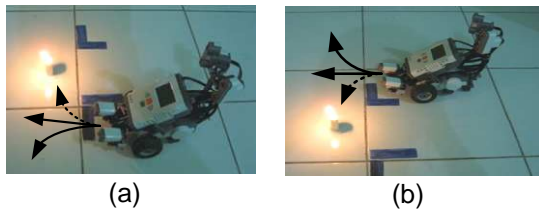


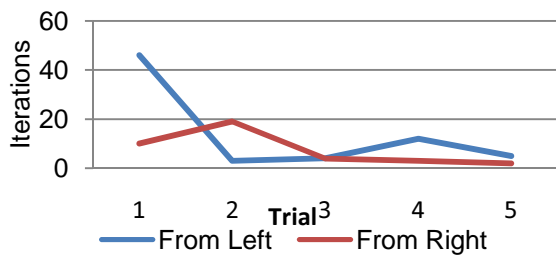Figure 20. Robot's performance by using search target behavior



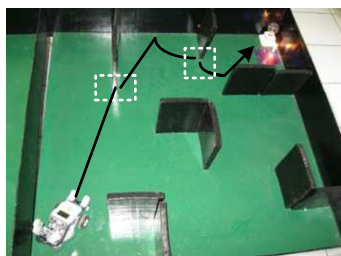Figure 21. Robot's performance on reaching the target



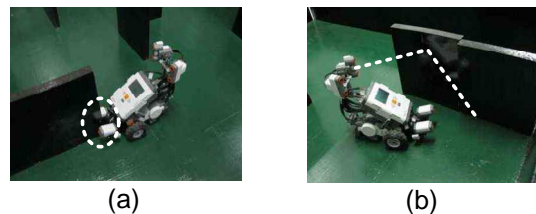Figure 22. Autonomous navigation of robot with Q learning - obstacle aviodance behavior



Figure 23. Problems on physical robots

## 5. Conclusion

It can be concluded that physical robot using subsumption architecture and motor schema as behavior coordination methods can finish navigation task well. Motor schema tend to give faster result on reaching target. It is happened because motor schema has more accurate (also slower) movement. However, subsumption architecture still has advantage on its robust (also faster) result and simple implementation.

Robot using Q learning mechanism can learn obstacle avoidance task well, this is remarked by its success in collecting positive rewards continually. Learning rate affect the robot's learning performance. When it is getting bigger, learning phase getting faster too. Although Q learning can be applied in search target behavior, but it does not give satisfying result in amount of positive rewards collected by robot. Hence it is suggested to be applied only on obstacle avoidance behavior. Physical robot applying Q learning can solve navigation task well, but there also weaknesses on light sensor placement and ultrasonic sensor characteristic.

## References
[1] Nillson NJ. *Shakey the Robot.* AI Center. SRI International. Technical Note 323. 1984.
[2] Brooks R. A Robust Layered Control System For a Mobile Robot. *IEEE Journal of Robotics and Automation*. 1986; 2(1): 14 – 23.
[3] Arkin RC. *Motor Schema Based Navigation for a Mobile Robot : an Approach to Programming by Behavior*. IEEE Int. Conf. on Robotics and Automation. 1987: 264-271.
[4] Arkin RC. Behavior-Based Robotics. England: Bradford Books. 1998.
[5] Pirjanian P. *Behavior coordination mechanisms: State-of-the-art*. Univ. Southern California. Technical Report IRIS-99-375. 1999.
[6] Sutton RS, Barto AG. Reinforcement Learning: an Introduction. Massachusets: MIT Press. 1998.
[7] Erden MS, Leblebicoglu K. Free gait generation with reinforcement learning for a six-legged robot. *Robotics and Autonomous Systems*. 2008; 56 : pp.199–212.
[8] Kroemer OB, Detry R, Piater J, Peters J. Combining active learning and reactive control for robot grasping. *Robotics and Autonomous Systems*. 2010; 58(10) : pp.1105–1116.
[9] Watkins C, Dayan P. Q-learning: Technical Note. *Machine Learning*. 1992; 8: 279-292.
[10] Perez MC. A Proposal of Behavior Based Control Architecture with Reinforcement Learning for an Autonomous Underwater Robot. Ph.D.Thesis. Girona: University of Girona; 2003.
[11] Wicaksono H, Prihastono, Anam K, Kuswadi K, Effendie R, Jazidie A, Sulistijono IA, Sampei M. *Modified Fuzzy Behavior Coordination for Autonomous Mobile Robot Navigation System*. Proc. of ICCAS-SICE. Fukuoka. 2009.
[12] Anam K., Kuswadi S. *Behavior Based Control and Fuzzy Q-Learning For Autonomous Mobile Robot Navigation*. Proceeding of The 4th International Conference on Information & Communication Technology and Systems (ICTS). Surabaya. 2008.
[13] Knudosn M, Tumer K. Adaptive navigation for autonomous robots. *Robotics and Autonomous Systems*. 2011; 59 : 410-420.
[14] Hohl L, Tellez R, Michel O, Ijspeert AJ. Aibo andWebots: Simulation, wireless remote control and controller transfer. *Robotics and Autonomous Systems*. 2006; 54(6): 472–485.
[15] Oliveira G, Silva R, Lira T, Reis LP. *Environment Mapping using the Lego Mindstorms NXT and leJOS NXJ*. EPIA. 2009.
[16] Benedettelli D., Ceccarelli N., Garulli A., Giannitrapani A. Experimental validation of collective circular motion for nonholonomic multi-vehicle systems. *Robotics and Autonomous Systems*. 2010; 58(8): 1028-1036.
[17] Adriansyah A. Sebuah Model Berbasis Pengetahuan untuk Pengendalian Formasi Sistem Robot Majemuk. TELKOMNIKA Indonesian Journal of Electrical Engineering. 2010; 8(2): 81–86.
[19] Djajadi A, Laoda F, Rusyadi R, Prajogo T, Sinaga M. A Model Vision of Sorting System Application Using Robotic Manipulator. TELKOMNIKA Indonesian Journal of Electrical Engineering. 2010; 8(2): 137–148.
[20] Leffler BR, Mansley CR, Littman ML. *Efficient Learning of Dynamics Models using Terrain Classification*. Proceedings of the International Workshop on Evolutionary and Reinforcement Learning for Autonomous Robot Systems. 2008.