Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1-3, 2008

Quantifying Coherence when Learning Behaviors via Teleoperation

Sekou Remy and Ayanna M. Howard

***Abstract*— Applications of robotics are quickly changing. Just as computer use evolved from research purposes to everyday functions, applications of robotics are making a transition to mainstream usage. With this change in applications comes a change in the user base of robotics, and there is a pronounced move to reduce the complexity of robotic control. The move to reduce complexity is linked to the separation of the role of robot designer and robot operator.**

For many target applications, the operator of the robot needs to be able to correct and augment its capabilities. One method to enable this is learning from human data, which has already been successfully applied to robotics. We assert that this learning process is only viable when the demonstrated human behavior is coherent. In this work we test the hypothesis that quantifying the coherence in the provided instruction can provide useful information about the progress of the learning process.

We discuss results from the application of this method to reactive behaviors. Such behaviors permit the learning process to be computationally tractable in real-time. These results support the hypothesis that coherence is important for this type of learning and also show that this property can be used to provide an avenue for self regulation of the learning process.

I. Introduction

The description of the average robot user is changing. Just as computers have evolved from the realm of research and extreme novelty applications to becoming commonly found in multiple places in a typical modern home, as well as in the modern workplace - robotics is also making a similar transition. Changing applications results in a change in the typical user and whether an amputee, an arthritic grandmother, or a health care worker providing in-home assistance, a general theme is that people seek to have less computational effort expended to control the devices they use. Also, as the user base becomes more non-technical, there is an increased push to reduce the need to be technically involved in robotic instruction.

With the change in demographics of the average robot user, there is a need to separate the role of robot designer and robot user. Such a separation would require that the user be able to provide new or fine tune existing robotic capabilities, and learning by teleoperation is one such method of accomplishing this task. For the changing user to utilize an in-home robot, we believe that this will allow the typical home user the ability to train the robot without placing heavy requirements on expertise and prior training. Learning from teleoperation can also be classified with similar approaches such as learning by demonstration [1], [2], and learning by observation [3], [4]. These methods have been applied to robotics and serve as a valid method of enabling a robot to perform new tasks.

Sekou Remy and Ayanna M. Howard are with the Human-Automation Systems (HumAnS) Lab, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta GA 30332, USA
{sekou, ayanna}@ece.gatech.edu

One challenge created by using such an approach is that the robot will need to be endowed with the ability to regulate when it learns in concert with its user/teacher/operator. We propose that this self regulation, a critical component of autonomous robotic learning from human supervision is hinged on some base properties that must be properly explored. One such property, coherence, has not been fully treated in robotics and in this work we seek to investigate coherence and its relationship to learning from human data.

II. Background

There are two challenges with incorporating home users into the robotic learning process. First, the user may not be able to identify when the limits of the robot's capabilities are reached and second, the user may not initially be capable of providing competent training to the robot. We believe that coherence is a property that can be used to address these two challenges.

Coherence is a term that has been mentioned in several engineering and science domains. Merriam-Webster defines coherence as a "systematic or logical connection or consistency", but in each domain the definition varies to some degree. In robotics the term has been used in [5], [6], [7] but even in these usages, a process to quantify coherence is not evident.

In this work we use the definition that coherence in teleoperation data is the property that forces the action state to be linked to a specific sensory state for each behavior. Coherence rises from a causal relationship between the actions executed and the sensory evidence provided. Because of this relationship the actions executed should be *logically consistent* with the evidence provided. In this work, we present an approach which enables a robot to quantify coherence in the instruction which it is provided. For humans, it is often times obvious that if data are not coherent this will pose challenges for learning. We believe that if robots are so equipped to also identify this property, then it can have positive impact, especially for applications in which autonomous robot learning is useful. The approach utilizes a property known as the mean quantization error which will be defined more fully in section III. By isolating characteristics of this property we show how it can be used to augment the process of learning from teleoperation when possible and also to identify when it is not possible for the robot to learn in that manner.

978-1-4244-2213-5/08/$25.00 ©2008 IEEE

III. METHODOLOGY

There are several possible methods that have been implemented to accomplish learning from teleoperation. This paper does not aim to present a comparative analysis of such methods, but it seeks to present a study on how coherence affected one of these methods [8]. This method uses an interactive learning method to incrementally learn a target behavior via teleoperation starting with zero initial knowledge. In this work we test the hypothesis that quantifying the coherence in the provided instruction can provide useful information about the progress of the learning process. This information is directly related to the performance of the learned behavior. We also seek to allow a robot to generate this information as the behavior is learned.

A. Behavior

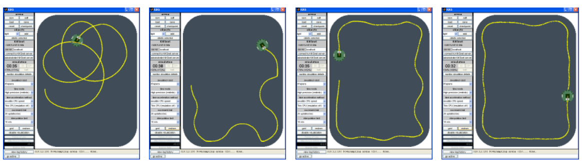
A behavior is defined as a specific relationship between evidence provided and actions executed. It is a policy of action that defines how the robot operates when specific sensor data is detected. If this relationship is causal, it implies that the behaviors under consideration are reactive. Mathematically a behavior can be represented as a mapping $f : X \rightarrow Y$ where X is the sensor space provided to the user for teleoperation, and Y is the action space of the haptic device used to capture human action. If the behavior is defined by teleoperation, the mapping is essentially shown by example. A given example (x_k, y_k) indicates that the action $y_k \in Y$ is executed when the sensor data $x_k \in X$ is detected. The example is an "if-then" rule that captures a facet of the behavior.

To highlight this process, Fig. 1 shows paths demonstrated at different stages as the robot learned. At the beginning of the interaction, there are few examples of the target behavior provided for the robot to learn from. As more examples are presented for learning, the robot is better able to learn the task.

B. Learning the Behavior

As outlined, teleoperation can be used to capture a set of examples performed by the human user. This set, defined in (1), provides examples of a behavior f_N. Through interactive learning, the target behavior is defined as more examples are provided for the learning process (as N increases). In essence, $f_N \rightarrow f$ if the learning process is successful.

$$B = \{(x_k, y_k)\}, \quad k \in \{1, ..., N\} \tag{1}$$



(a) 10 (b) 25 (c) 150 (d) 400

Fig. 1. Navigation behavior learned as more human examples are provided.

where N is the number of training points provided.

The interactive learning approach utilized also incorporated two Self Organizing Maps [9] to reduce the dimensionality of the sensing and the acting spaces by mapping $X \rightarrow \hat{X}$ and $Y \rightarrow \hat{Y}$. This neural network inspired tool was utilized so that the process of determining the behavior f would become tractable in realtime and without bias for the target behavior or robotic platform. Using SOMs permitted the sensing and acting state to be adaptively classified - with more examples the classifiers became better tuned to the input data provided. The learned behavior thus becomes $\hat{f_N} : \hat{X} \rightarrow \hat{Y}$. If the learning process is successful, $\hat{f_N} \rightarrow f$ as N increases.

C. Coherence

Coherence is a property that describes the logical consistency between sensing and action for the behavior demonstrated via teleoperation. If the user is demonstrating a coherent behavior, they consistently perform the same actions when presented with the same sensor input. The use of a classification method to identify sensor and action data, especially an adaptive method, lies as the core of quantifying coherence.

The **mean quantization error** (MQE) of the classifier is a value which represents how accurately the method is able to classify data. This term is defined in (2) where u is the number of classes, n is the dimension of the data, and m_i is the best matching unit (BMU) in the classifier for the data item a_k. a_k is either x_k or y_k depending on whether the MQE is being calculated for sensing or actuation data.

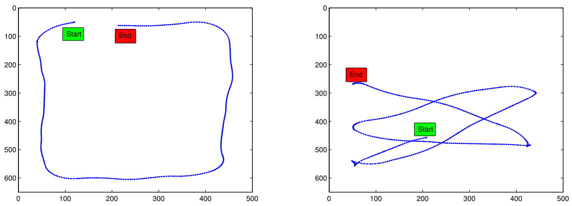
$$MQE = \frac{1}{u} \sum_k \frac{1}{n} ||m_i - a_k|| \tag{2}$$

D. Characteristics of the MQE

The quantization error of an adaptive classifier can be used to provide insight to how the classified data changes over time. As each new data item is presented, the classifier will generate MQE values that increase. Once the data values presented are no longer novel, the MQE will increase less since the data items will more closely match their BMU.

It is important to note that only in cases where the data are almost the same as the classifier's BMUs will the MQE values decrease. This means that if the classifier is starting as a blank slate with no knowledge of the data, initially the MQE should increase rapidly, since every new data item will be poorly classified. As more data items are presented it is likely that some of them will be similar to items previously presented, if there is coherence in the demonstrated behavior. As such, after an initial adaptation period where the MQE increases, it can be expected that the MQE will stop increasing. If learning is successful, the rate at which examples are provided should reduce as well after some initial learning period (although these two periods might not fully coincide).

If the robot is not able to learn the task, the reduction in intervention - where the human provides examples of the target behavior - will not occur. If this inability to learn is



(a) Wall Following. (b) Wandering.

Fig. 2. Path of teacher demonstrating behavior.

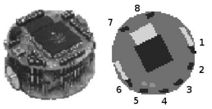


Fig. 3. Real and simulated Khepera robot. Sensors 1 to 8 are labeled.

related to poor classification capabilities, the MQE values generated by the classifier will also increase, although possibly at a slower rate than the initial adaptation period. Poor classification could either be the result of inadequacies in the sensors provided, in the fidelity of the actuator provided, or even through short comings in the classification method itself. This understanding is filtered down into (3) which quantifies coherence in order to predict performance after period j.

$$\text{prediction}_j = \sum_{p=Beg_j}^{End_j - 1} max\{MQE_{p+1} - MQE_p, 0\} \quad (3)$$

Where Beg_j and $End_j - 1$ mark the first and the penultimate examples during period j. MQE_p is the error calculated after the p^{th} example has been provided.

IV. Experimental Setup

A. Scenario

For this work, two behaviors were selected for study. The first, wall following is a staple in many navigation applications. The second, wandering, fills a useful niche in “fetch and carry” applications. This behavior was also selected since it was a sufficiently distinct behavior from wall following. Both of these are among the five behaviors indicated as basis behaviors in [10]. Figs. 2(a) and 2(b) show an example of each behavior.

Simulation was used to provide a level of control over environmental factors. The simulated robotic platform used was based on the Khepera mobile robot (K-TEAM) and endowed with eight infrared sensors and two actuators providing differential drive (see Fig. 3). The simulated sensors possessed a noise profile in line with the hardware sensors after which they were fashioned.

B. Evaluating Coherence

To artificially inhibit the ability to demonstrate coherent behavior by teleoperation, noise was introduced into sensing

TABLE I

Cases considered

actuator\sensor	not modified	modified
not modified	Case 1	Case 2
modified	Case 3	Case 4

and action. This noise, or random data, reduced the consistency in human action. The effect of modifying coherence was evaluated by attempting interactive learning under each of the cases in Table I, then observing the outcomes in each situation. For each form of the cases, the target behavior remained the same - wall following.

To introduce noise in the sensing process (as in Cases 2 & 4), sensor data from sensors 4,5 and 6 (see Fig. 3) were replaced by three independently identically distributed uniform random numbers $U_i \sim U(0, 1024)$. These sensor values were critical to the task of wall following. Since U_i is not related to actual sensor values, the new values are effectively “decoupled” from the actual sensors. The sensor state $\hat{X}$ used in the learning process is a corrupted version of the actual sensor state.

To introduce noise in actuation (as in Cases 3 & 4), values provided through the use of the haptic device were modified by adding two normal random numbers $N_i \sim N(1, 5)$.

As an aside, it is useful to note that introducing randomness is useful in another important way as well. Other than providing an artificial method to reduce coherence in a controlled manner, randomness can exist as a natural feature of teleoperation. The robot’s ability to identify when its learning is impeded by intrinsic randomness is also of great value.

C. Experiments

The experiments in this work used human subjects with varying degrees of expertise to teach a mobile robot interactively. Each subject performed the target behavior teleoperatively for one lap then entered into an interactive learning phase where the robot learned for ten laps. During interactive learning the user incrementally provided examples of the target behavior and observed the robot’s actions. Whenever examples were provided to the robot it incorporated them into its behavior and then demonstrated the updated behavior. Through this incremental process the user and the robot simultaneously adapted to each other. A two-axis joystick was used to capture human action and instead of providing just sensor values to the human, the overhead view of the robot in the simulated arena was provided (see Fig. 1).

D. Evaluating Performance

To evaluate the performance of the wall following behavior, the examples gathered during interaction were grouped into laps. To calculate the performance after the k^{th} lap, the examples provided during laps $1, ..., k$ are collected and used for learning. This is one advantage of the learning approach used in that it is able to apply data interactively or in batch form.

The metric used to compute performance is presented in (4). It is a two part construct of the time to complete a lap, t and the distance to the nearest wall, d.

$$\text{performance} = \alpha_1 d + \alpha_2 t \tag{4}$$

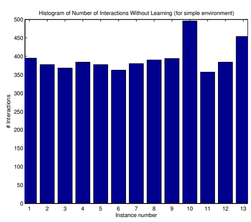
The weights α_1 and α_2 are extracted from the values of $1/d$ and $1/t$ when the behavior was executed using a controller devised using evolutionary techniques. By definition the performance of this controller is 2 and for the listed performance metric, *smaller* values are produced with improved performance. It is noted that expert human operators can demonstrate better performance than this coded controller, but the purpose of using its weights in this manner is to provide a basis for comparison.

V. RESULTS

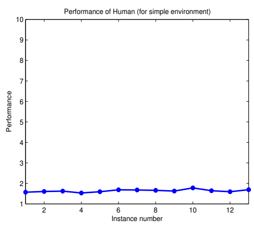
To provide a baseline for comparison, results are first presented for teleoperation without the effect of learning. Fig. 4 shows the number of interventions required for the robot to travel a single lap while performing the wall following behavior. When learning was introduced, the number of interventions required decreased significantly for the first lap and then decreased further as interaction time increased in each case (as shown in Fig. 5). Since interactive learning was applied, situations where the robot learned to perform the appropriate task reduced the need for the operator to intervene. Cases where the levels of intervention decreased over time indicate cases in which the operator determined that the robot was learning the task. Such a pronouncement is confirmed by the improvements in performance over time that are also presented in Fig. 6. In all considered cases some level of learning was attained.

The levels of interaction needed for Cases 2 & 4, are higher than for Cases 1 or 3, yet the performance demonstrated in the former cases is poorer. This indicates that not only was learning less successful in Cases 2 & 4 but that they also required more input from the operator.

To get a better viewpoint on the success of learning, the mean quantization plots are presented. These plots (Fig. 7) show the MQE for both sensor and actuation classifiers in each of the four cases. As expected, each plot can be considered in two phases. The first phase occurs primarily during the first lap of interactive learning where the classifiers learn

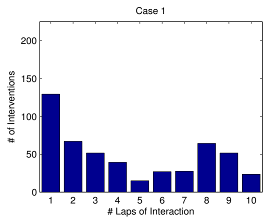


(a) Number Interactions.

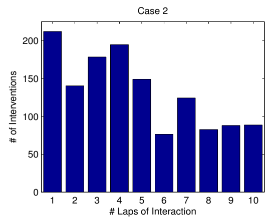


(b) Performance.

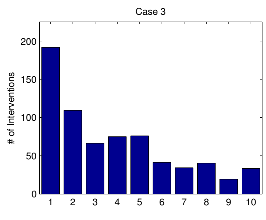
Fig. 4. Teleoperation (without learning).



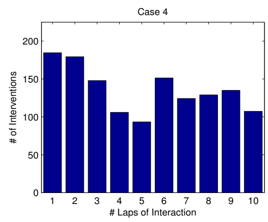
(a) Case 1.



(b) Case 2.



(c) Case 3.

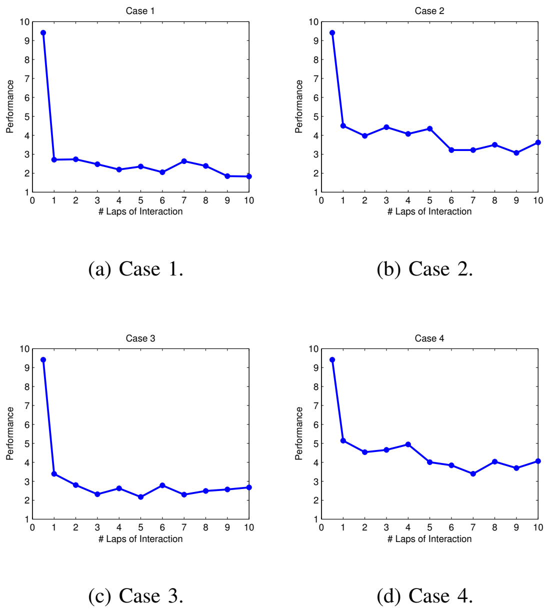


(d) Case 4.

Fig. 5. Interaction level vs. # laps.

to classify examples of sensor and actuation data. During this phase the largest increases in MQE are found since in each case the learning begins from zero initial knowledge. The second phase occurs over the remaining nine laps. In cases where classifiers are presented with perturbed data, Figs. 7(c), 7(f), 7(g) & 7(g), the MQE values increase steadily over this phase. The MQE values in the second phase of these cases are also significantly higher than cases where like classifiers (either sensing or actuation) are considered. This confirms the theory that the classifiers are impeded by incoherence caused by randomness. Further support is drawn from the observation that the level of interaction over this phase does not significantly drop in these incoherent cases. In the case where randomness was introduced only in sensing (Case 2), the MQE plot for the MQE values for the actuation classifier do not significantly increase (see Fig. 7(d)) however the interaction level remains relatively constant. The same can be seen in Fig. 7(e) which shows the MQE for the sensing classifier in the case where randomness was introduced in the actuation (Case 3). These two plots show a situation that indicates the classifiers are neither getting better nor getting worse. Were these plots not coupled with MQE plots with values that increased, they would have identified the case where the user thought robot learning was inadequate yet the classifiers had reached their limits. Such cases will be investigated more fully in future work.

Fig. 8 shows the paths demonstrated by robots that learned in each of the four cases. All robots were taught the same wall following task but only in Case 1, where coherence was not violated, does the robot perform the task as well as the operator who trained it. These results, especially those presented in Fig. 7, show that coherence has a measurable effect on learning from teleoperation. This measure of co-



(a) Case 1. (b) Case 2.

(c) Case 3. (d) Case 4.

Fig. 6. Performance vs. # laps.

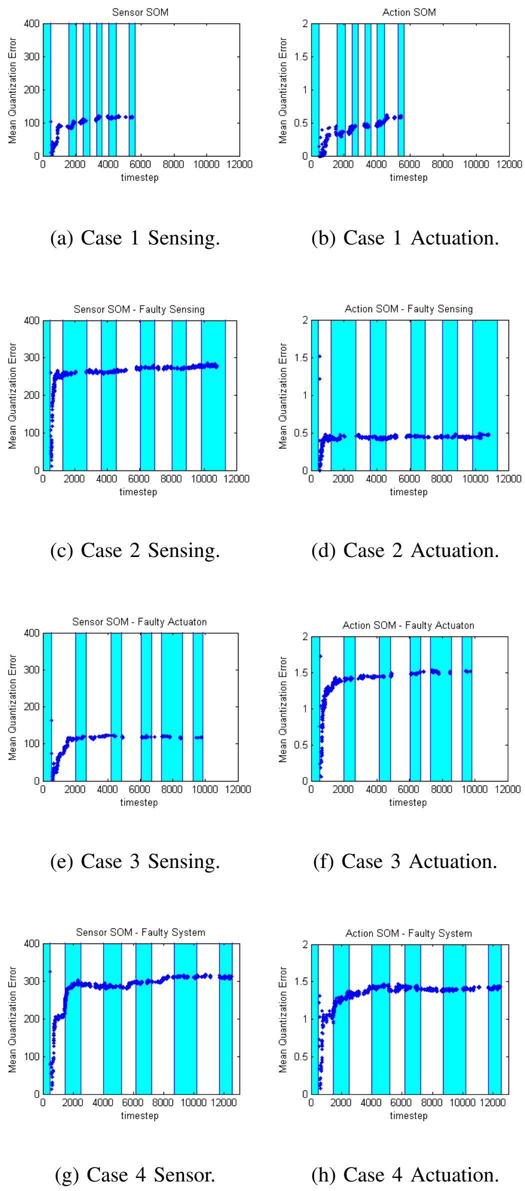
TABLE II
EFFECT OF APPLYING STOPPING RULE

Case	#laps	Performance	%Performance lost	%Examples saved
1	6	2.02	9.18	50.8
2	3	4.43	18.2	151.0
3	3	2.32	-15.6	87.0
4	5	4.01	-1.42	91.0

herence can be used to predict robot performance during the learning process. This effect, quantified in (3), permits the graphs shown in Fig. 9 to be generated. As anticipated, the relationship found in these plots closely mirrors that presented in Fig. 6 (measured performance). As an example of how coherence be used to regulate learning, a simple stopping rule is evaluated: Stop learning at lap $j > 1$ if $\text{prediction}_j - \text{prediction}_{j+1} < 0$. The results of this rule are presented in Tab. II. These results show that the amount of time required for human instruction is significantly reduced, while still maintaining satisfactory robot performance. Because of space limitations the results presented focus heavily on only one of the two target behaviors. The second, wall avoidance, while a vastly different behavior also demonstrated similar results. Fig. 10(a) and 10(b) show the MQE plots for example of wall avoidance. Many of the properties mentioned for wall following are also present in these graphs. Fig. 10(c) shows the path of a robot that learned the behavior and Fig. 10(d) shows the predicted performance plot. This predicted performance plot is quite significant in this case since a suitable performance metric for this behavior was difficult to define.

VI. Summary and Future Work

While it is a foregone conclusion to some that learning from incoherent data is not ideal, it is an important area of



(a) Case 1 Sensing. (b) Case 1 Actuation.

(c) Case 2 Sensing. (d) Case 2 Actuation.

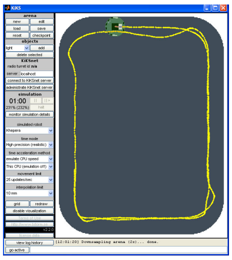
(e) Case 3 Sensing. (f) Case 3 Actuation.

(g) Case 4 Sensor. (h) Case 4 Actuation.

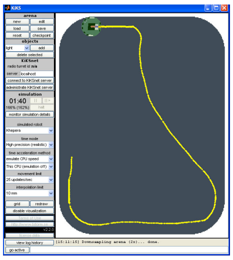
Fig. 7. MQE vs. Time (superimposed on lap boundaries).

study since endowing the robot to identify such situations has not yet been presented. In this work incoherence was introduced to behaviors by artificially adding randomness and the effect on learning from teleoperation was studied. It was shown that there was a measurable difference in performance, but there was also a systematic change in the properties of the quantization error. Capturing this change enabled coherence to be quantized using data that can be extracted as instruction is occurring.

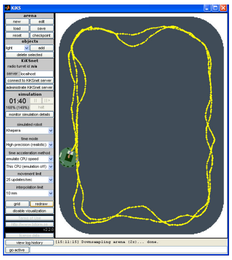
One application of the quantification of coherence was presented which indicated the savings in time and effort that could have been provided given this information. In general, such information can be directly used to autonomously regulate the learning process and also to enable the robot to provide useful feedback to the user. Both of these are



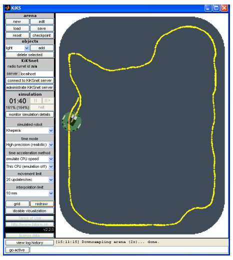
(a) Case 1.



(b) Case 2.

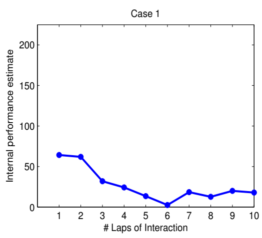


(c) Case 3.

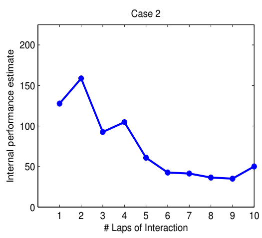


(d) Case 4.

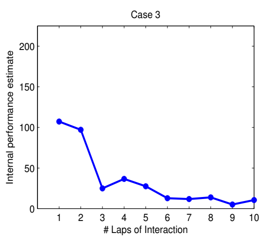
Fig. 8. Robot demonstrating learned behaviors.



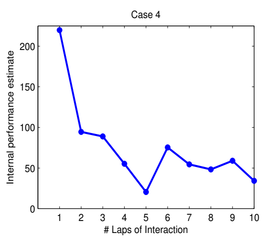
(a) Case 1.



(b) Case 2.

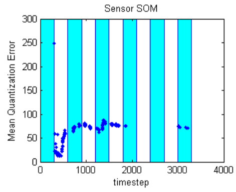


(c) Case 3.

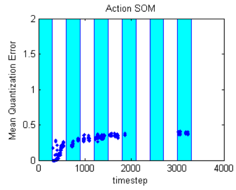


(d) Case 4.

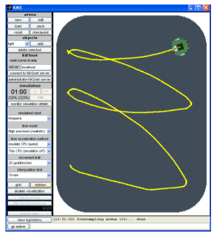
Fig. 9. Predicted Performance vs. # laps.



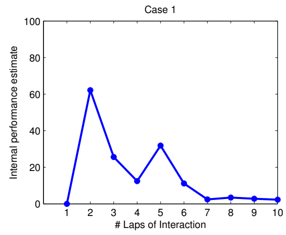
(a) Case 1 Sensing.



(b) Case 1 Actuation.



(c) Learned path.



(d) Predicted performance.

Fig. 10. Wall avoidance behavior.

capabilities that would well serve users who teach robots by teleoperation. This work thus provides useful results that will help cope with the future needs of the in-home robot user.

VII. Acknowledgements

This research is based upon work supported by the National Science Foundation under Grant No. IIS-0705130.

References

[1] C. G. Atkeson and S. Schaal, "Robot learning from demonstration," in *Proc. 14th International Conference on Machine Learning*. Morgan Kaufmann, 1997, pp. 12–20. [Online]. Available: citeseer.ist.psu.edu/atkeson97robot.html

[2] D. H. Grollman and O. C. Jenkins, "Dogged learning for robots," in *2007 IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2007.

[3] M. N. Nicolescu and M. J. Mataric, "Natural methods for robot task learning: instructive demonstrations, generalization and practice," in *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems*. New York, NY, USA: ACM Press, 2003, pp. 241–248.

[4] D. C. Bentivegna, A. Ude, C. G. Atkeson, and G. Cheng, "Learning to act from observation and practice," *International Journal of Humanoid Robotics*, vol. 1, no. 4, pp. 585–611, June 2004.

[5] C. Breazeal, "Learning by scaffolding," Ph.D. dissertation, M.I.T. Department of Electrical Engineering and Computer Science, 1998. [Online]. Available: citeseer.ist.psu.edu/breazeal98learning.html

[6] R. A. Brooks, "Integrated systems based on behaviors," *SIGART Bulletin*, vol. 2, no. 4, pp. 46–50, 1991.

[7] K. Noda, M. Suzuki, N. Tsuchiya, Y. Suga, T. Ogata, and S. Sugano, "Robust Modeling of Dynamic Environment Based on Robot Embodiment," in *IEEE International Conference on Robotics and Automation (ICRA'2003)*, 2003, pp. 3565–3570. [Online]. Available: http://icra.ntu.edu.tw/icra/

[8] S. Remy and A. M. Howard, "In situ interactive teaching of trustworthy robotic assistants," in *Proc. IEEE Conference on Systems, Man, and Cybernetics*, Montreal, Canada, Oct. 2007.

[9] T. Kohonen, *Self-Organizing Maps*, 3rd ed., ser. Springer Series in Information Sciences. Springer Verlag, 2001, vol. 30, ISBN 3–540–67921–9, ISSN 0720–678X.

[10] M. J. Mataric, "Getting humanoids to move and imitate," *IEEE Intelligent Systems*, vol. 15, no. 4, pp. 18–24, 2000.