Eye tracking with Educational Robots: A Cautionary Tale

ABSTRACT

We present an eye-tracking study on an educational robot Thymio II with 52 participants. The participants observed the robot detecting obstacles in the first phase of the experiment and they interacted with the robot in the second phase of the experiment. Finally, they were asked to explain the functionality of the robot. The gaze of the participants was recorded during both the phases. The values from the robot's sensors were visualized on the robot's body using LEDs in two different ways in the two experimental conditions and there was no sensor data visualization in the control condition. We found that the sensor data visualization has an impact on the gaze patterns and the performance of the participants. The main goal of the experiment was to investigate that eye tracking could bring insights to the effectiveness of the interaction between the human user and the robot.

1. INTRODUCTION

Eye tracking provides unprecedented access to the users' attention and engagement during interactive scenarios. Previous research has shown that gaze data can be used as a proxy to understand the underlying socio-cognitive aspects in not only human-computer interaction but in human-human interaction as well. In the present decade, off the shelf eye-trackers have become readily available to the researchers to assess the users' attention and engagement.

On the other hand, educational robots can be of great use in making young school kids interested in technological developments [1, 17]. It is important to know about the key socio-cognitive features of this subset of human-robot interaction because these features let the researchers design better interactive scenarios in the future. These features include the conceptual mapping of the sensor data visualization and the robot's function and the generalizing the concepts to solve a given problem in a contextual learning environment.

Thymio II [10] can provide salient effects onto our visual attention through the LEDs. We can use this feature to guide the users' attention in a way that the effectiveness of the interaction is maximized. In the case of interaction with the educational robots the effect of the interaction is to make the users understand the technical aspects of the robot and to make the users be able to transfer it to different domains.

We present a study with combining eye tracking and human robot interaction (HRI) in an educational setting. Eye tracking data is of high temporal precision as compared to some of the physiological data sources such as fMRI or other biometric devices used to measure the palpitation and the body temperature. The eye tracking devices are portable enough as opposed to the bulky apparatus used in fMRI [11]. Eye tracking data provides more direct access to the users' attention than other portable and physiological data sources such as EEG; this makes eye-tracking data a pertinent source to measure attention.

Evaluation of HRI is mostly done through qualitative questionnaires and qualitative data (interaction videos, interviews for example). The physiological measurements such as eye tracking being precise, pertinent and portable at the same time provides an added value to the researcher. However, this needs to be complemented with the qualitative data as well.

We propose to use the mobile eye trackers to be able to have our participants freely move on the experiment site. Using mobile eye tracker poses a technical challenge to automatically locate where the participants are looking. This is not a trivial task as the visual stimulus for participants is changing with every bit of motion of their head. This is the reason we used fiducial-markers on the maze we used in the experiment so that we can reconstruct the whole maze even if only a part of it is the visual field of the participant.

Our working hypothesis is that with careful mapping between the LED visualizations and the behavior of the robot we can induce the understanding of the functionality of the robot very easily. We propose to use the gaze data to verify this hypothesis. With eye-tracking data we can judge whether the mapping between the behavior and visual actuators is useful or not for the participants' understanding. Moreover, we hypothesize that the good performers will be able to anticipate the robot's motion. Eye-tracking data can also be useful to verify the anticipation hypothesis.

The experiment consists in two phases. The first phase is the observation phase during which the participants watch the robot move and detect an obstacle in a "playground". The second phase requires the participants to interact with the robot by placing an obstacle in its path to make it de-

10th ACM/IEEE International Conference on Human-Robot Interaction 2015 Portland, USA

tect the obstacle before reaching a certain target. After each phase the participant explain the behavior of the robot. The sensor data was visualized through LEDs with three visualization conditions: TRUE visualization, RANDOM visualization and NONE visualization. Through this experiment we investigate the following research questions:

1. Are there different gaze patterns for different levels of performance during the interaction? Moreover, can we see any anticipation patterns for the good performers?

2. How does displaying the sensor information on the robot' body as a visual cue affects the gaze and performance of the participants?

The rest of the paper is organized as follows. The second section gives a brief review of the eye-tracking research done to quantify the expertise and performance. The third section presents the specificities of the study. The fourth section lists the main results. The fifth section discusses the results. Finally, the sixth section concludes the paper.

2. RELATED WORK: GAZE PATTERNS AND EXPERTISE/TASK BASED PERFORMANCE

Several scholars related gaze patterns with level of expertise. [4] in an air traffic-monitoring task found out that the experts looked less at the scenario-specific information than novices. [3, 7, 16] studied the effect of expertise on the gaze patterns in different surgical tasks and concluded that experts look less at the instruments than the novices, instead they focus more on the task specific areas. [9] showed that expert chess players pay more attention to the relative positions of the pieces, rather than the individual pieces, than novice chess players. [**?**] also studied the difference between experts and novice chess players in a checkmate avoidance task and concluded that the experts have more gaze falling on the important squares than the novices. In a program-debugging task, [12] showed that the expert programmers scan through all the lines in the program faster than the novices. In a collaborative Tetris game, [6] showed that experts pay more attention to the stack of Tetronimoes while novices allocate more attention to the new pieces falling from the top.

Our previous work also show a clear relation between gaze patterns and task-based performance. In a pair-programming task, [5] showed that the good performing pairs have more synchronized gaze on different parts of a program than the bad performing pairs. In a similar task, [13] showed that the good performing pairs pay more attention to the data-flow of the program than the poor performing pairs. Moreover, [14] showed that while describing the functionality of a program the well performing teams had more gaze on the variable modification parts in the program while poor performing teams have equal distribution of gaze on different parts of the program during similar phase of the task. Thus we see that previous research provides insights about the relationship between the gaze patterns and the behavioral and task-based performance indicators in diverse scenarios.

It must be noted that to our best knowledge only very few studies have investigated how eye-tracking could be used to assess HRI. In an eye-tracking study, in the context of human robot interaction, [15] found that when the robot provides a matching visual and verbal reference to an object the participants were more likely to look at the correct objects. This result was consistent irrespectively of the ambiguity of the situation.

Figure 1: **Thymio II**: the proximity sensors were used to detect the obstacle based on the reflectance of the material. The sensor data was shown as a proportion of the circle lit using the LEDs on top of the robot.

3. EXPERIMENT

3.1 Setup and Conditions

Thymio II is a small robot designed for education for 6 to 16 year old children [8, 10]. Users can interact with it via buttons and distance sensors. The robot can display its states via many LEDs and its motion. We programmed it to show the value of the front IR-sensor on its top. The display consists of 8 LEDs (Figure 1).

In the experiment, there were 3 sensor data visualization conditions. For the TRUE visualization condition, the number of illuminated LEDs, on the top of the robot, is proportional to the intensity measured by the IR-sensor. When the intensity of the reflected IR crosses a predefined threshold, the robot does a 90° turn to its right. For the two other conditions, the behavior is the same, simply the display changes. In the RANDOM condition, the display shows as if the robot was detecting the obstacle at random moments, while in the NONE visualization condition nothing is shown on top of the robot. The robot has a fitting for a pencil that we used for the drawing of its trajectory.

The playground (Figure 2) is designed in order to allow the participants to get a reference of the previous behavior of the robot, as well as to allow us to localize the position of the robot, the position of the obstacles and the position of the gaze on the video recorded from the eye-tracker. The localization is made possible by the use of unique fiducial markers on the playground. On the right hand side, the observation phase takes place and there will be the reference for the black and white obstacles. On the left hand side, the interaction takes place.

One important fact to be considered is that the localization of the robot, the obstacle or the gaze pointer on the observation field for the participant is not a trivial task. The fact that the participants are free to move, poses a challenge to the eye-tracking analysis. The movement of the participant causes change in the observation field of the participant almost every frame of the eye-tracking video. To cope up with the changing observation field we decided to put an

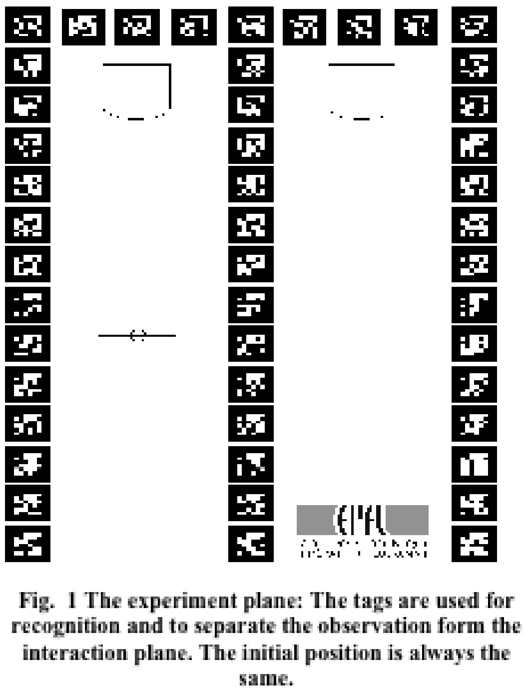

Figure 2: Basic playground for the experiment. The fiducial markers are used for automatic localization of the gaze pointers on the observation video.

array of fiducial markers on the playground. Using the fiducial markers enabled us to recreate the whole playground for every frame in the video recorded from the eye-tracker's camera (Figure 3).

3.2 Procedure

Upon their arrival at the experiment site, the participants signed a consent form and answered a small questionnaire about demographics and participant's experience with educational robots (this we later consider as expertise). The participants were placed in front of the playground (Figure 3) and were equipped with the SMI eye-tracking glasses. The actual experiment consisted in two phases: observation phase and interaction phase.

Observation phase.

The participant observes the Thymio II robot approaching obstacles. The robot turns as soon as it detects the obstacle. The participant watches the robot's behavior for a black and a white obstacle for 5 trials each. For the white, the robot turns earlier; for the black obstacle the robot turns later. In the final trial the robot is equipped with a pen to draw its trajectory for both obstacles with different colors. After drawing, the participant is asked the question: "How does the robot work?"

Interaction phase.

In the second phase, the participant is asked to guide the robot to a goal, using a gray obstacle. The participant can put the obstacle wherever he likes on the left hand side (Figure 2) of the playground before the robot starts. We designed the gray obstacle to reflect more infrared light (IR) than the black or the white obstacles therefore the robot turns even earlier. This leads to a surprising effect. The participant has only the robot's trajectories as a source of information. Each participant is given 5 trials to put the obstacle in order to guide the robot to the goal. The participant then has to answer the same question as in the observation phase, response to which is considered as a final answer.

3.3 Participants

We recruited 52 participants for the experiment. All of them were students from École Polytechnique Fédérale de Lausanne, Switzerland. Most of them were in their first year undergraduate program (with two exceptions). The age distribution was narrow, with a mean of 19.7 and a standard deviation of 1.97. There were 13 female and 39 male participants. In the three experimental conditions, NONE, RANDOM and TRUE, there were 18, 17 and 17 participants respectively.

3.4 Performance measures

We had two measures of performance. First measure was a rating given to the responses that participants gave to the question about the functionality of the robot. Second measure was the distance from the target in the interaction phase.

We transcribed the responses to perform a rating scheme on the level of abstraction and correctness. A low level of abstraction consists in describing the functionality of the robot in the terms of the electric or programming aspects of the robot, a high level represents describing the behavior.

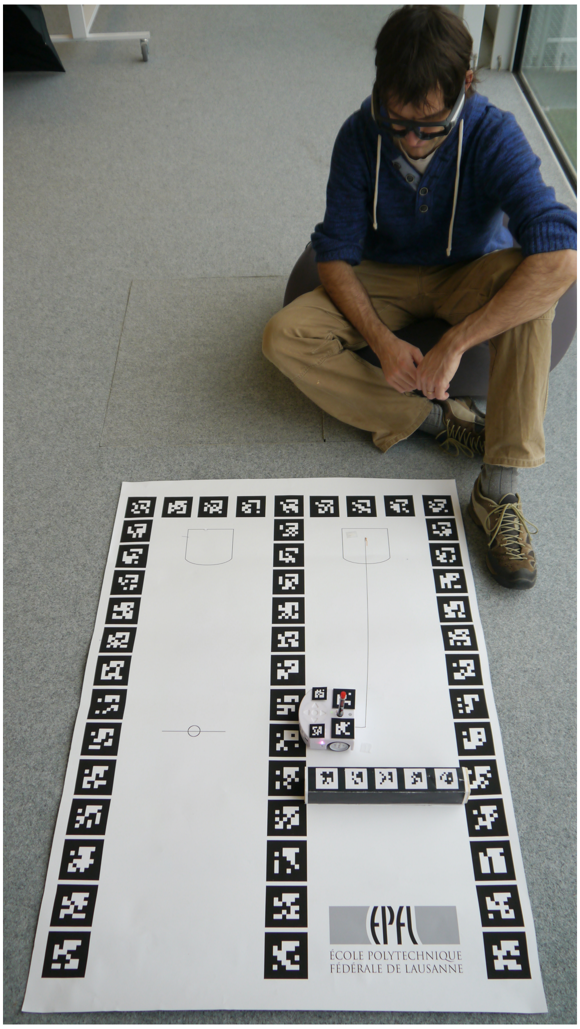

Figure 3: Experimental setup for observation phase. The observation video is recorded from the video camer in the front of the eye tracker (in the lop-left corner).

For the levels of correctness, we categorized the responses into correct or incorrect response. In TABLE I. we present sample responses for the different categories.

Correct.	Abstract.	Sample answer
incorrect	*low*	“The robot turns at a given distance in respect to a color, defined in the program.”
incorrect	*high*	‘Gray is the most dangerous, so the robot turn really early.”
correct	*low*	“The gray obstacle reflects a lot, the robot sends light. At a certain threshold the robot turns.”
correct	*high*	“The gray obstacle is shinier, so more reflective, and better detectable.”

Table 1: Sample answers for all classes

Second, the distances from the target were measured for all 5 trials. As a matter of fact, as soon as the principle of the robot was deemed to be understood by the participants, almost all the participants behaved the same way. On the first try almost every participant failed and the robot turned instantly. So the most significant distance measure was the improvement between the first and second trial. It is referred to as improvement in the following sections.

Incorrect answers were further classified into 4 categories; we present each category accompanied by a sample response:

- The participant did not respond or offer a possible explanation: “I really don’t know how the robot works”.
- The participant identified a wrong cause for the behavior of the robot: “The robot notices the tags on the side”.
- The participant held the code responsible for the cause of different behavior: “The robot turns at a given distance in respect to a color, defined in the program”.
- The participant tried to explain by analogies or did not explain the robots behavior at an understandable level: “Gray is the most dangerous, so the robot turns really early”.

3.5 Gaze measures

The most common form of eye-tracking data aggregation are the fixations. Fixations depict the periods in the user’s observation when the user attends a relatively small part of the visual stimulus for a relatively longer period of time. The various fixation detection algorithms are summarized in [2]. Following are the two gaze measures, derived from the fixations, which we used in the present work.

Average fixation duration on the robot: We measured the amount of time a participant looks at the robot and averaged this duration for the number of fixations on the robot for the particular participant.

Average fixation on the reference side: The main area (right hand side of figure 2) in the observation phase is treated as the reference side in the interaction phase. The rationale behind keeping the lines drawn by the robot in the observation phase is for the participants to be able to refer to the prior behavior of the robot in the interaction phase.

We measured the amount of time a participant looks at the reference side and averaged this duration for the number of fixations on the reference side for the particular participant.

4. RESULTS

We found no bias for age, gender, expertise, student status or condition in respect to the correctness, abstraction and type of mistake made for the final answer. Surprisingly there is no correlation between the visualization condition and the correctness or abstraction of the answers given. We could not find any anticipation patterns in the eye-tracking data. We found some other relation between the different measures. We present the results in following subsections.

4.1 Average fixation duration on the robot during observation phase vs. condition

The average fixation duration on the robot during observation phase (Figure 4) is significantly more in TRUE and RANDOM condition than in the NONE condition ($F[2,49] = 3.68$, p = .03). This depicts that the sensor data visualization has an effect on the users' attention.

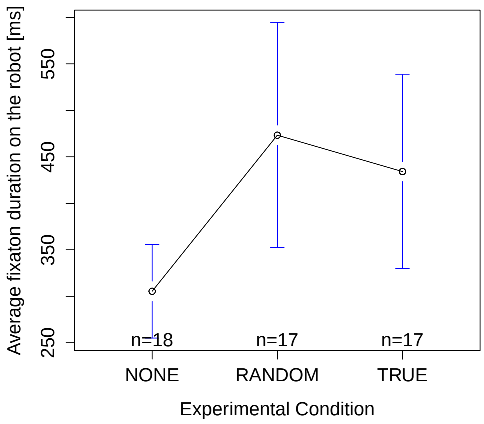

Figure 4: Average fixation duration on the robot during the observation phase vs. Condition

4.2 First improvement vs. condition

The first improvement is significantly more in NONE and RANDOM conditions (Figure 5) than in the TRUE condition ($F[2,49] = 3.75$, p = 0.03).

4.3 Average fixation duration on the reference side during interaction phase vs. condition

The average fixation time on the reference side of the playground (Figure 6) during the interaction phase is significantly more in NONE and RANDOM condition than it is in the TRUE condition ($F[2,49] = 4.19$, p = .02).

4.4 First improvement vs number of fixations on reference side

There is a significant negative correlation between the number of fixations on the reference side (Figure 7) during the interaction phase and the first Improvement ($t(50) = -2.13$, Pearson's correlation = -0.29, p=.03).

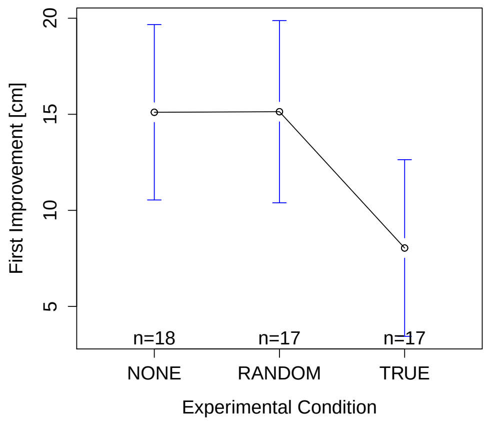

Figure 5: First improvement vs. experimental condition

Figure 6: Average fixation duration on the reference side during the interaction phase vs. Condition

5. DISCUSSION

We present an eye-tracking study in the context of human-robot interaction with an educational setting. The fact that participants have higher average fixation duration on the robot during the observation phase in the NONE condition than the other two conditions (Figure 4) is not surprising as displaying visual information on the robot induces a saliency effect on the attention and the gaze is attracted towards the salient feature in the field of view.

The fact that the participants in the RANDOM or NONE visualization condition improved more than those with TRUE condition is surprising (Figure 5). There are two plausible explanations. The first is that participants who see the robot in the TRUE condition have a stronger believe that the robot always behaves the same. They see the LEDs of the robot turning on and therefore the robot still works. The ones with NONE visualization do not have an indication whether the robot works the same way as in the observation phase, and start experimenting earlier. The second is that the par-

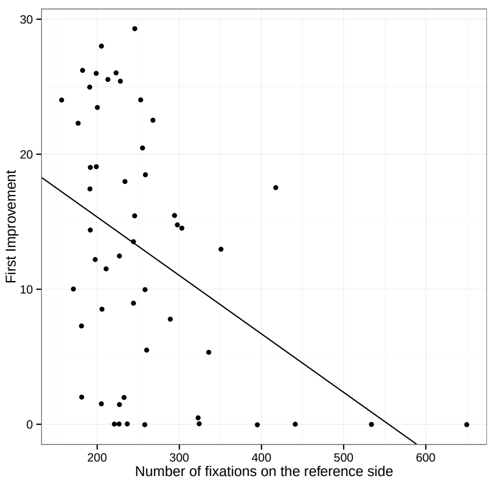

Figure 7: First improvement in centimeters (y-axis) vs. number of fixations on reference side (x-axis) during the interaction phase.

ticipants in TRUE condition just put the obstacles out of detection range, because they could see different display information and concluded therefore it may behave differently.

The fact that participants have lower average fixation duration on the reference side in the interaction phase in the TRUE condition than the other two conditions actually supports the explanation that participants with TRUE condition had a higher initial belief in their cognitive model of how does the robot work (Figure 6). They did not need to look onto the reference given in observation phase. This also verifies our working hypothesis that a carefully designed mapping between the robot's behavior and its visual actuators can help in maximizing the learning effect during the interaction.

We demonstrated that eye-tracking data can provide several process variables to assess not only the task-based performance of the participants but also to evaluate HRI. The prime advantage of using mobile eye-tracking is that the users' interaction with the robot can still be ecologically valid. On the other hand, using fMRI for evaluating HRI reduces the interaction to a video of the robot only [11].

Another advantage of using mobile eye-tracking to evaluate HRI, especially in an educational context, is that the data provides direct and noise free access to the users' attention. While in other physiological measures for attention, such as EEG, the data has a lot of noise and the relation to the attention is very task-specific.

The advantages of mobile eye-tracking come with an additional cost. As we mentioned earlier, the automatic localization of the gaze pointer on the visual stimulus is not a trivial task. To do this we have to use a number of fiducial markers to the playground (Figure 2). We observed that the fiducial markers cause a distraction to the participants. As a future work, we are trying to find a suitable replacement to the fiducial markers.

Another plausible explanation for the participants in NONE and RANDOM condition being better performers than the TRUE condition during the interaction phase comes from the transparency of the robot's behavior. The robot's behavior was not at all transparent in the former two conditions. This might have caused the participants to put more effort in understanding the behavior. Therefore, they performed better.

6. CONCLUSION

Our effort to use the confluent of eye tracking and human-robot interaction provides insights to two aspects: first, as a way to evaluate the human-robot interaction and second, as a way to evaluate the interaction scenario.

We addressed two research questions in the beginning of the present work. The results show that the way the sensor data is visualized has an impact on both the gaze patterns and the performance (Question 2). However, for the relation between the gaze patterns of the participants and their performance outcomes (Question 1) we cannot claim a direct relation. We found a negative correlation between the gaze-patterns and the first improvement during the interaction phase. We could not find a relation between the explanation of the participants and their gaze-patterns. Moreover, we did not find any anticipation patterns in the gaze data. The lack of evidence for the relation between participants' explanations and their gaze patterns prevents us to make any strong claims about the relation between the overall performance and the gaze patterns.

The fact that we cannot find a relation between the gaze patterns and all the factors in the performance might have its roots in the way we designed the interaction phase experiment. As a working hypothesis we said that there should be clear and carefully designed mapping between the robots' behavior and the sensor data visualization. Its environment also affects the behavior of the robot and hence the environment also should be taken into consideration while designing the behavior-visualization mapping. We introduced a novelty (the gray obstacle being more reflective) in the environment during the interaction phase of the experiment that forced the participants to revise their hypothesis and hence they failed to understand the robot's functionality. This also serves as a word of caution in future while designing such interactions. The observation phase serves as a training session for the participants, which can later be tested during the interaction phase. Any novel behavior modality (in our case, the more reflective gray obstacle) can have a negative impact on the participants' understanding.

In a nutshell, the use of gaze data as an assessment of the human-robot interaction in an educational setting provides future research prospects. The data acquisition is a better trade-off between the ecological validity of the experiment (as there cannot be an interaction with the robot fMRI data acquisition) and the precision of the data stream (wristbands for measuring the palpitation and the temperature).

Acknowledgments

This research was supported by...

7. REFERENCES

[1] M. Cooper, D. Keating, W. Harwin, and K. Dautenhahn. Robots in the classroom-tools for accessible education. *Assistive technology on the threshold of the new millennium*, pages 448–452, 1999.

[2] A. Duchowski. *Eye tracking methodology: Theory and practice*, volume 373. Springer, 2007.

[3] S. Eivazi, R. Bednarik, M. Tukiainen, M. von und zu Fraunberg, V. Leinonen, and J. E. Jääskeläinen. Gaze behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 377–380. ACM, 2012.

[4] C. Hasse, D. Grasshoff, and C. Bruder. How to measure monitoring performance of pilots and air traffic controllers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 409–412. ACM, 2012.

[5] P. Jermann and M.-A. Nüssli. Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1125–1134. ACM, 2012.

[6] P. Jermann, M.-A. Nüssli, and W. Li. Using dual eye-tracking to unveil coordination and expertise in collaborative tetris. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*, pages 36–44. British Computer Society, 2010.

[7] B. Law, M. S. Atkins, A. E. Kirkpatrick, and A. J. Lomax. Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 41–48. ACM, 2004.

[8] S. Magnenat, F. Riedo, M. Bonani, and F. Mondada. A programming workshop using the robot "thymio ii": The effect on the understanding by children. In *Advanced Robotics and its Social Impacts (ARSO), 2012 IEEE Workshop on*, pages 24–29. IEEE, 2012.

[9] E. M. Reingold, N. Charness, M. Pomplun, and D. M. Stampe. Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, 12(1):48–55, 2001.

[10] F. Riedo, P. Rétornaz, L. Bergeron, N. Nyffeler, and F. Mondada. A two years informal learning experience using the thymio robot. In *Advances in Autonomous Mini Robots*, pages 37–48. Springer, 2012.

[11] A. M. Rosenthal-von der Pütten, F. P. Schulte, S. C. Eimler, L. Hoffmann, S. Sobieraj, S. Maderwald, N. C. Krämer, and M. Brand. Neural correlates of empathy towards robots. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 215–216. IEEE Press, 2013.

[12] B. Sharif, M. Falcone, and J. I. Maletic. An eye-tracking study on the role of scan time in finding source code defects. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 381–384. ACM, 2012.

[13] K. Sharma, P. Jermann, M.-A. Nüssli, and P. Dillenbourg. Gaze evidence for different activities in program understanding. In *24th Annual conference of Psychology of Programming Interest Group*, number EPFL-CONF-184006, 2012.

[14] K. Sharma, P. Jermann, M.-A. Nüssli, and P. Dillenbourg. Understanding collaborative program comprehension: Interlacing gaze and dialogues. In *Computer Supported Collaborative Learning (CSCL 2013)*, number EPFL-CONF-184007, 2013.

[15] M. Staudte and M. W. Crocker. Visual attention in spoken human-robot interaction. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 77–84. ACM, 2009.

[16] G. Tien, M. S. Atkins, B. Zheng, and C. Swindells. Measuring situation awareness of surgeons in laparoscopic training. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 149–152. ACM, 2010.

[17] F. Wyffels, M. Hermans, and B. Schrauwen. Building robots as a tool to motivate students into an engineering education. *AT&P JOURNAL PLUS*, 2:113–116, 2010.