# Using mediator objects to easily and robustly teach visual objects to a robot.

Pierre Rouanet, Pierre-Yves Oudeyer, David Filliat

# Using mediator objects to easily and robustly teach visual objects to a robot.

Pierre Rouanet, Pierre-Yves Oudeyer and David Filliat

Social robots are drawing an increasing interest both in scientific and economic communities and one of the main issues is the need to provide these robots with the ability to interact easily and naturally with humans. We believe that the interaction issues may have a very strong impact on the whole system and should be given more attention. Current research however focus mainly on the the visual perception and/or machine learning issues (see for example Steels and Kaplan [1]). We think that by focusing on the users and on the interface we can help them provide the learning system with very high quality learning examples. In particular, we think that we should focus on the following questions:
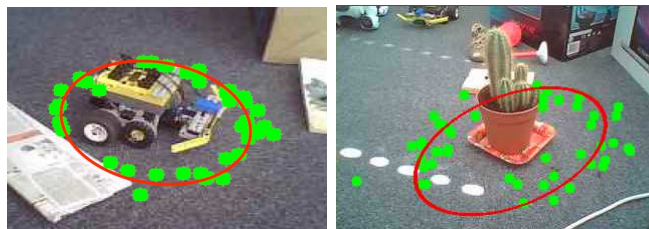
Yet, we think that by focusing on the users and on the interface we can help them to provide the learning system with really good quality learning examples. In particular, we think that we should focus on the following questions:

**Attention drawing:** How can a human smoothly and intuitively draw the robot's attention toward the interaction?
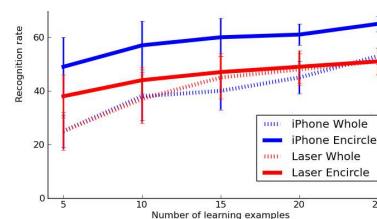
**Pointing and Joint attention:** How can a human robustly designate an object to the robot? How can a human understand what the robot is paying attention to?

One can try to address these challenges by transposing the human-like interactions, such as gaze tracking or pointing gestures. However, most social robots have a visual apparatus and in particular a small field of view which makes this kind of interaction non-robust and very restrictive in real environments. Other researchers directly wave objects in front of the camera of the robot and so can achieve a motion-based joint attention [2]. Although this approach is interesting it can only work with light and movable objects and therefore could be tiring or even impossible for the elderly or the disabled. We are here proposing to use small devices as mediator objects between the human and the robot. We already presented an iPhone based interface [3] and we are here presenting a Wiimote and laser pointer based interface. This interface allows users to drive the robot and to draw its attention toward a specific object in order to name it. The laser spot is automatically tracked by the robot and a laser sound is played as a visual feedback allowing users to know whereas the laser spot and so the designated object was inside the robot's field of view. It is a crucial help as non-expert humans have difficulties to correctly estimate the robot's capacities. To name an object, users have to first encircle it. On top of being a simple gesture to select an object, it is also providing a rough visual segmentation which is otherwise still an ill-defined problem in an unconstrained environment.

We are also presenting an evaluation of the laser interface and its comparison with other interfaces and especially the



iPhone interface mentioned above. We design an experiment where we ask participants to teach the robot names for five different objects. Participants used three different interfaces: the first one was very simple and did not provide any feedback of what the robot sees, the second was the laser interface and the third was the iPhone interface. We first study the overall quality of the learning examples among these different interfaces. We noticed that the object was entirely visible on only 25% of the learning examples gathered without any feedback and even entirely absent in 37% of these examples. While the object was entirely visible in more than 90% of the learning examples collected with an interface providing a feedback (laser or iPhone). We also used the gathered learning examples to train a learning system based on the bag of visual words and evaluate its performance in generalization on an offline database. We showed that, while the laser interface allows the user to provide high quality examples, encircling with the laser is not as effective as encircling on the screen of the iPhone. Indeed, the projection of the laser spot on the camera plane often results in cutting the encircled object as shown in the pictures above.

## REFERENCES

[1] L. Steels and F. Kaplan, "Aibo's first words: The social learning of language and meaning," *Evolution of Communication*, vol. 4, no. 1, pp. 3–32, 2000. [Online]. Available: http://www3.isrl.uiuc.edu/ jun-wang4/langev/localcopy/pdf/steels02aiboFirst.pdf

[2] F. Lömker and G. Sagerer, "A multimodal system for object learning," in *Proceedings of the 24th DAGM Symposium on Pattern Recognition*. London, UK: Springer-Verlag, 2002, pp. 490–497.

[3] P. Rouanet, P.-Y. Oudeyer, and D. Filliat, "An integrated system for teaching new visually grounded words to a robot for non-expert users using a mobile device," in *Proceedings of the Humanoids 2009 Conference*, 2009.