



**HAL**  
open science

# A Robotic Game to Evaluate Interfaces used to Show and Teach Visual Objects to a Robot in Real World Condition

Pierre Rouanet, Fabien Danieau, Pierre-Yves Oudeyer

► **To cite this version:**

Pierre Rouanet, Fabien Danieau, Pierre-Yves Oudeyer. A Robotic Game to Evaluate Interfaces used to Show and Teach Visual Objects to a Robot in Real World Condition. ACM/IEEE Human-Robot Interaction HRI 2011 conference, Mar 2011, Lausanne, Switzerland. hal-00647459

**HAL Id: hal-00647459**

**<https://inria.hal.science/hal-00647459v1>**

Submitted on 2 Dec 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Robotic Game to Evaluate Interfaces used to Show and Teach Visual Objects to a Robot in Real World Condition

Pierre Rouanet  
INRIA, France  
pierre.rouanet@inria.fr

Fabien Danieau  
INRIA, France  
fabien.danieau@inria.fr

Pierre-Yves Oudeyer  
INRIA, France  
pierre-  
yves.oudeyer@inria.fr

## ABSTRACT

In this paper, we present a real world user study of 4 interfaces designed to teach new visual objects to a social robot. This study was designed as a robotic game in order to maintain the user's motivation during the whole experiment. Among the 4 interfaces 3 were based on mediator objects such as an iPhone, a Wiimote and a laser pointer. They also provided the users with different kind of feedback of what the robot is perceiving. The fourth interface was a gesture based interface with a Wizard-of-Oz recognition system added to compare our mediator interfaces with a more natural interaction. Here, we specially studied the impact the interfaces have on the quality of the learning examples and the usability. We showed that providing non-expert users with a feedback of what the robot is perceiving is needed if one is interested in robust interaction. In particular, the iPhone interface allowed non-expert users to provide better learning examples due to its whole visual feedback. Furthermore, we also studied the user's gaming experience and found that in spite of its lower usability, the gestures interface was stated as entertaining as the other interfaces and increases the user's feeling of cooperating with the robot. Thus, we argue that this kind of interface could be well-suited for robotic game.

## Categories and Subject Descriptors

I.2 [Artificial Intelligence]: Robotics

## General Terms

Human Factors

## Keywords

HRI, language teaching, social robotics, interface design

## 1. INTRODUCTION

Social robots are drawing an increasing amount of interest in the context of domestic services or entertainment. So, it is

crucial to provide these robots with the ability to interact in novel and changing environments. Among the many issues that this involves, we are here interested by the problem of how a non-expert human can show and name a new visual object to a robot such that it can recognize it later on.

We previously proposed a state-of-the-art computer vision and machine learning framework, based on an incremental version of the visual bags of words approach [3], which can be used to robustly categorize visual objects when trained with "good enough" learning examples [12]. However, the performance in generalization of such a system is extremely dependent of the quality of the learning examples given as an input [12]. Therefore, if one is interested in solving this problem in a real human-robot social interaction (HRI) context, we need to first answer the questions: how to collect good learning examples through few and intuitive interactions with non-expert humans? And, how can we collect examples by using current social robots which have typically limited sensors and in particular a strongly constrained visual apparatus? Among the many challenges raised by these questions, we are focusing on: **Attention drawing**: How can we easily and robustly draw the robot's attention toward ourself or toward the interaction? **Pointing**: How can a human point and draw the robot's attention toward a specific object? How can the robot separate the object from its background in a cluttered environment if it does not already know the object and thus does not already have a model of the object? **Joint attention**: How can the user be sure of what the robot is paying attention to? And how can the robot be sure of what the human is paying attention to? [1][6]

In previous work, we introduced several human-robot interfaces based on mediator objects specifically designed to address these challenges and improve the efficiency of our learning system. For instance, we already presented a Wiimote interface, a laser interface and an iPhone interface [10][11][12]. These interfaces allow intuitive controls and different kinds feedback about what the robot is perceiving to help users collect good examples.

So far, these interfaces were only evaluated in small-scale artificial pilot studies with mostly robot-expert participants [10][11][12]. Yet, as explained above, the long term goal of this project is to allow real users, i.e. non-expert users, to easily and robustly teach visual objects to their social robots. Indeed, it should allow them to help their robot discover their environment, in a way which has similarities with way parents help their child discover objects around them. Therefore, there is a real need of conducting real

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HRI'11, March 6–9, 2011, Lausanne, Switzerland.

Copyright 2011 ACM 978-1-4503-0561-7/11/03 ...\$10.00.

world study of this kind of interaction to make sure that potential users, correctly understand the task and intuitively, easily and efficiently manage to have the robot collect good learning examples. Interfaces also have to be non-restrictive and even entertaining in order to keep the users interested, especially if one envisions lifelong interactions.

Thus, in this paper, we present a real world user study which takes place in a sciences museum in Bordeaux from June to November 2010 (see figure 1). In this experiment, we ask 107 naive participants to show and teach names for visual objects to a social robot. This study was designed as a robotic game in order to embed this a-priori artificial task in a more realistic scenario and so maintain the user's motivation during the whole experiment. The participants used one of the above mentioned interfaces (Wiimote, laser and iPhone) or an hand gestures interface based on a Wizard-of-Oz (WOZ) recognition setup. This fourth interface was added in order to compare the mediator based interfaces to an a priori more natural and direct kind interaction, which as we will show reveals to be less usable and less efficient than the mediator based interfaces.



**Figure 1:** The real world study designed as a robotic game to investigate how users teach a social robot names for new visual objects. In particular, we studied the impact of the interface on the quality of learning examples gathered.

More precisely we try to answer the following questions:

- How each given interface impacts the learning examples quality? And in particular, can non-expert users provide good learning examples without any feedback on what the robot is perceiving?
- Are our interfaces easy-to-use? Intuitive? Entertaining? Do the different interfaces significantly impact the user's experience?

Next to this central questions for assessing our interfaces, the very concept and instantiation of the robotic game also raises interesting questions, especially since robotic games is a research area still mainly unexplored [2][18]: Is the game entertaining and immersive? How easily can one learn how to play with it?

## 2. RELATED WORK

Many researchers have studied how a human can teach new visually grounded words to a robot. For instance, Steels and Kaplan developed a complete social framework based on human-like interactions such as pointing gestures and speech recognition to allow users to teach words to an AIBO robot [15]. In this work, the authors were not specifically focusing

on the interface issues. Yet, the HRI was identified has a major limitation of their system. In particular, they showed that the lack of robustness of the interface often leads to some bad learning examples and so decreased the overall learning performance [5]. Scassellati developed mechanism of shared attention with a humanoid robot through gaze monitoring and pointing gestures [14]. However, in this work, he used a fixed upper-torso and thus restricted the interaction.

As shown in the examples mentioned above, most authors choose to use human-like interactions to try to address the HRI challenges. This approach potentially provides really natural interactions. Unfortunately, existing associated techniques, for gesture and gaze recognition and interpretation in particular, are not robust enough in uncontrolled environments (due to noise, lighting or occlusion) and most social robots have a body whose shape and perceptual apparatus is not compatible with those modes of interaction (small angle of view, small height...). Yet, it is really important that users can robustly designate objects in order to avoid misunderstanding or frustration during the interaction but also to help users collecting good learning examples. Indeed, in a cluttered environment a non-robust pointing may lead to designate the wrong object and thus a completely wrong learning examples which will decrease the whole learning system performance (see figure 2).



**Figure 2:** To allow users to designate a particular object to a robot in a cluttered environment, we need to provide them with a robust and accurate pointing detection. Otherwise it may lead to a restrictive interaction and even to false learning examples.

One way to circumvent this pointing and joint attention problem, is to allow users to directly wave objects in front of the camera of the robot [8][17]. However, users can only show to the robot small and light objects which can be easily carried. Thus, we can not show objects such as a table, a plug or a painting on a wall. Moreover, for the elderly or the disabled waving objects could be really tiring or even impossible.

We argue that one way to help to achieve intuitively and robustly some of the functionalities presented above, without facing the problems encountered when waving objects, is to develop simple artifacts that will serve as mediators between the human and the robot to enable natural communication, in much the same way as icon based artifacts were developed for leveraging natural linguistic communication between human and certain bonobos [13]. Interfaces based on mediator objects have already largely been used in HRI and especially to draw a robot's attention toward an object. For instance, Kemp et al. or used a laser pointer to

easily and robustly designate objects to a robot in order to ask it to fetch them [7]. Here, they used the laser pointer as a point-and-click interface. They showed that naive participants managed to correctly designate objects to a robot. Ishii et al. proposed a laser pointer-based interface where users can draw stroke gestures using the laser to specify various commands such as path definition or objects selection with lasso gestures [4]. However, in their work they used calibrated ceiling-mounted cameras and vision-based ID tags to circumvent the object recognition issues.

Even though, many interfaces used to show and/or teach visual objects to a robot have been developed, few real world studies have been designed to evaluate how non-expert users really interact with such interfaces. This is what we achieve in this article in order to assess the interfaces introduced in [10][11][12].

### 3. OUTLINE OF THE SYSTEM

As stated above, we developed a full system, i.e. visual recognition, machine learning and human-robot interaction. However, as the purpose of this paper is centered on the user study of the different interfaces, the first two parts of our system will not be described here (details can be found in [3] and [12]). In this section, we will only describe the four interfaces used for the study. They were chosen to span the variety of mediator interfaces that one can imagine but also to explore the different kind of feedback of what the robot is perceiving that we can provide to the users. Each of these interfaces has to provide the user with the abilities to **drive the robot and draw its attention toward a direction or an object**.

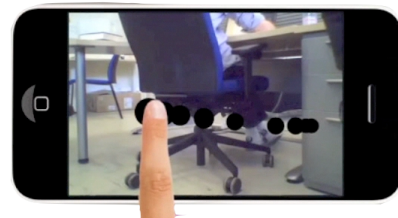
The interface was not used to trigger the capture of a new learning example, i.e. when users think that the robot sees the object they want to teach. Instead, they had to directly touch the head of the robot. We choose this interaction to avoid the feeling of tele-operating the robot and thus increase the user's feeling of cooperating with it.

It is important to notice here that all the interfaces were based on the exact same sensorimotor capacities and functionalities of the Nao robot. This choice seems interesting to us as the Nao sensorimotor apparatus represents well the present of existing social robots. Thus, we voluntarily choose not to enhance its capacities by using ceiling or wide range camera. Indeed, our goal was to study how a non-expert user can show and teach visual objects to an actual social robot.

#### 3.1 iPhone interface

This interface is based on an iPhone. The video stream of the robot camera is displayed on the screen, as shown on figure 3. It allows users to accurately monitor what the robot really sees. This is a key feature to achieve joint attention situation. However, the user's attention is split into direct and indirect robot's monitoring which may lead to increase the user's cognitive workload.

The touchscreen is also used as an interactive gesture board. For instance, users can draw strokes directly on the visual feedback such as a vertical stroke from the bottom to the top of the screen to make the robot move forward. They can also tap on a particular spot to ask the robot to aim its head on this specific location, which is very practical way to draw the robot's attention toward an object.



**Figure 3:** The iPhone interface with the video stream, allowing the users to monitor what the robot is looking at. Users can also define “trajectories” on the touch-screen (through strokes such as swipe or tap) interpreted as commands.

#### 3.2 Wiimote interface

This interface is based on a Wiimote device (see figure 4). The users can press one of the button of the directional cross to move the robot. For instance, if you press the up button the robot will go forward until you release it. To aim the head of the robot, users have to orient the Wiimote (i.e. we directly map the values of the accelerometers to the pan/tilt values of the robot's head). Users can always focus their attention on the robot.



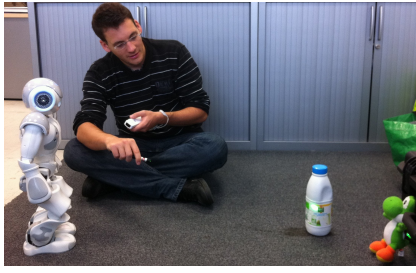
**Figure 4:** Users can move the robot by using the directional cross or directly orient the Wiimote to aim its head. However, the lack of feedback make it very difficult to estimate whether the robot really sees the object the user wants to teach it.

This interface does not provide any feedback about what the robot is perceiving, so the user has to “guess” what the robot really sees.

#### 3.3 Wiimote and laser interface

As in the previous interface, the Wiimote is used to drive the robot. However, as shown on the figure 5, in this interface a laser pointer is combined to the Wiimote and used to draw the robot's attention. Indeed, the robot is automatically tracking the laser spot and aims its head in order to keep it near the center of its sight. Here, users can draw the robot's attention toward a direction by smoothly aiming its head toward the right direction or they can point to a particular object, once it is inside the robot field of view, directly by designating it with the laser pointer.

To help users better understand when the robot is detecting the laser spot and thus correctly estimate the field of view of the robot, we provide them with an haptic feedback by rumbling the Wiimote each time the laser spot was detected by the robot. Yet, with this feedback users can not be



**Figure 5: Users can drive the robot with a Wiimote. They can also draw its attention toward an object by pointing it with a laser pointer.**

sure that the robot entirely sees the object they are designating but they can only be sure that the robot is detecting the spot visible on a part of the object.

### 3.4 Gestures-based interface (with WOZ)

In this interface, users can guide the robot by making hand or arm gestures. As gesture recognition is still a hard task, we used a Wizard-of-Oz framework where a human was controlling the robot accordingly to the different gestures he recognized. Users could make any gestures they wanted. However, the wizard was only seeing the interaction through the robot’s eyes. Indeed, as stated above we did not want to enhance the robot’s capacities with a ceiling or wide-angle camera as we wanted to study a direct interaction between non-expert users and what we think represent well a typical actual social robot. Thus, all the four interfaces are based on the same robot sensorimotor capacities.



**Figure 6: In this mode of interaction, the robot is guided by the hand and arm gestures made by the user. In order to have a robust recognition, we used a WOZ framework, where the wizard was only seeing through the robot’s viewpoint.**

Obviously, this interface embeds a strong difference with the others. The Wizard is a human who already has a knowledge and strong biases about what may constitute an object shown to him, i.e. the object segmentation problem is here automatically solved by the human. Thus, when users were pointing toward an object, the wizard was naturally centering the sight of the robot on it. However, this interface is centrally interesting for us, since it allows to assess how far the human-robot interaction could go if one would assume human-level computer vision and interpretation capabilities within the constraints of a typical social robot sensorimotor apparatus. Furthermore, this provides an interesting com-

parison with mediator based interfaces that do not use a Wizard-of-Oz. It will allow us to verify the hypothesis that due to the constrained visual apparatus of such typical social robot (small height and a limited field of view), this kind of natural interaction may in fact lead to a more restrictive and thus less satisfying interaction than with the other interfaces, in spite of human-level vision processing and interpretation.

## 4. EXPERIMENTS

As explained previously, we wanted here to study how real users manage to draw the attention of a robot toward an object present in its close environment and then associate a name with it. In particular, we try to study the impact that the interface may have on this task and on the interaction. We specially focus on:

- **Efficiency:** Study the impact of the interface on the quality of the learning examples gathered by participants and in particular the role of the feedback of what the robot is perceiving.
- **Usability and user’s experience:** Study how intuitive, effortless and entertaining our interface are. This is particularly important to encourage users to interact with their robot.

Potential future users of social robots will not necessarily be expert users, in the sense that they would like to be able to interact with their social robot without any prior specific training. Thus, it was crucial that our study was not a lab study but a real world study. In particular, we wanted to recreate a plausible context of interaction and have representative participants, so non-expert users, in order to preserve the ecological validity of our results and avoid the classical pitfalls in HRI [16][19].

However, as those users have probably never interacted with a social robot before, asking them to show and teach objects to a robot is still an unusual and artificial task as shown by pilot studies [10][11][12]. Therefore, we needed to embed it in a scenario in order to justify it. In particular, we needed to encourage the users to collect high quality learning examples. Moreover, we wanted a scenario that can entertain and maintain the user’s motivation during the whole experiment. In addition, we wanted to conduct a large scale study, and so we needed to design a formal and reproducible experiment.

### 4.1 A Robotic Game Experiment

We argue that one solution to tackle the above mentioned issues was to design our user study as a robotic game. Indeed, games are well known to be a powerful way to captivate and engage users, thanks to their scenario. For instance, serious games have been largely used for education or training, allowing learners to experience situations that are impossible in the real world [9]. Furthermore, we argue that, in the same way that video games have managed to make novice users solving complex and unusual tasks by using mechanism such as tutorial or briefing, we could design a robotic game experiment that lead users to achieve the complex task of teaching visual objects to a robot. The scenario of the game also allows us to justify this artificial task. Finally, presenting the experiment as a game allows us to attract a wide and various panel of participants. Moreover,



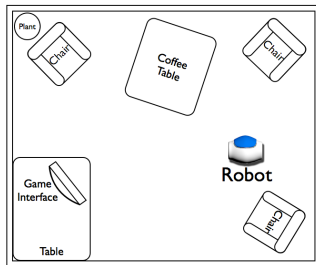
participants felt more comfortable participating a game than a scientific experiment.

Thus, we created a game scenario to try to match all the above requirements. The story was the following one: A robot came from another planet and was sent to Earth in order to better understand what seems to be a popular human habit: “playing soccer”. Indeed, from their remote home, the robots just picked up partial informations about this practice and so they want to investigate further. Therefore, the robot was sent to the living room of a soccer fan to gather more clues. As the robot was damaged during its journey, it can not fulfill its mission alone and so the user was asked to help the robot collecting these indications (i.e. collect learning examples of 4 different objects related to soccer). Every time the robot collected a new examples, a false and funny interpretation was given by the robot.

## 4.2 Experimental setup

### 4.2.1 Game Environment

We recreate a typical  $10m^2$  living room located next to the café of the museum (a map of the setup can be seen in the figure 7). We disposed furnitures such as tables or chairs and many various everyday objects (newspaper, plush, posters, etc...) in order to make it look inhabited. Among these various objects 12 were directly related to soccer (see figure 8). Those objects were the possible clues the robot needed to collect. They were chosen because they fit well within the scenario but also because they were textured and big enough so they can be robustly recognized by classical visual recognition algorithms.



**Figure 7:** We recreate a typical  $10m^2$  living room. To make it inhabited, furnitures were added along as well as many ornamental objects. Most of them were related to soccer in order to fit well within our scenario.

The design of the game environment had three main purposes:

- Reproduce a daily life area to provide participants with a stressless environment and to reduce the feeling of being evaluated.
- Conduct the experiment in a realistic environment, so users have to navigate the robot through a cluttered environment and to collect real world learning examples (lighting condition, complex background, ...).
- Immerse users into the scenario.

The global structure of the room remains unchanged during the whole experiment in order to get a constant test environment. Nevertheless, the small objects were randomly



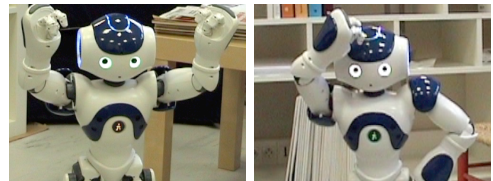
**Figure 8:** For the experiment, we used 12 textured objects directly related to soccer. Each participants had to teach 4 of these objects (randomly chosen among the 12 in each experiment) to the robot to help it better understand soccer.

disposed every 5 experiments. Indeed, in a real home, while big objects as furnitures will not move, most of the small objects will be often moved and thus must be recognized in spite of their background.

### 4.2.2 Robot

For this experiment, we used the Nao robot designed by the company Aldebaran Robotics<sup>1</sup>. Indeed, to us, it well represents the present of social robotics: with a toy-aspect and classical inputs (camera, microphone). As it is a humanoid, users will probably be more prone to teach it new words. But, as it is not too human-like it fits well within our alien robot scenario.

To make it more lively, we developed some basic behaviors such as yawning or scratching its head if the robot was idled for a long time. We also used different color for its eyes to express simple emotions or to provide feedback to the users (see figure 9). Moreover, we added organic sounds to express the robot mood.



**Figure 9:** Behaviors, such as “happy” on the left or scratching its head on the right, were designed to make the robot look more lively and help the users better understand its behavior.

Finally, as Nao is used for the soccer league of the RoboCup<sup>2</sup>, we can find many illustrations of the Nao playing soccer used to illustrate our scenario.

### 4.2.3 Game Interface

As explained above, to design our robotic game we draw our’s inspiration from classical video game. We used a large

<sup>1</sup><http://www.aldebaran-robotics.com/>

<sup>2</sup><http://www.robocup2010.org/>

screen as a game interface to display information to users such as cutscene video explaining the story.

This interface was also used to recreate a tutorial where participants learn one ability at a time: walking straight, turning, aiming the head of the robot and collecting a learning example. For each step, a short video explained how to realize the task with the interface they were using. After the video, the user was asked to effectively realize the task. Once it succeeds, it could move on to the next stage. These video were also a way to make users know better the robot and vice versa. Indeed, the final objective of the tutorial was to collect a first learning example which, in fact, was a picture of the user's face associated it with his name. The whole tutorial lasted about 5 minutes in average. After the tutorial, the real mission was explained to the participants thanks to another video such as the one on the figure 10. Thus, the game interface allowed us to present the whole experiment (both tutorial and mission parts) in one single game.



**Figure 10: The story of the game was told through video displayed on our game interface. This display was also used to provide users with step-by-step instruction of the tutorial.**

Furthermore, it also allowed us to conduct each test in the same way. Indeed, all participants received the exact same information and instructions through the game interface.

### 4.3 Experimental Protocol

The experiment took place from June to November 2010 and 107 persons participated to it. Most of them (74) were recruited at Cap Sciences, a science museum in Bordeaux. We expected there general non-expert public. Although, it might introduced a bias as sciences museum visitor are probably more receptive to technology. However, such bias are inevitable in HRI and was here clearly well identified. The other (33) were recruited on the campus of Bordeaux University of technology. We expected there participants with an important technological background and a knowledge of the classical interfaces but without any particular robotic knowledge.

77 participants were male, 30 were female and they were aged from 10 to 76 ( $M=26.3$ ,  $STD=14.8$ ). Among the 107 participants: 32 used the iPhone interface, 27 the Wiimote interface, 33 the Wiimote-laser interface and 15 the gestures interface.

Each participant was asked to follow the following protocol, validated by few pilot studies:

1. Fill in a consent form.
2. Fill in a pre-questionnaire.
3. Experimentation (robotic game)
  - Tutorial

- (a) Wake up the robot by touching its head.
- (b) Make it move forward.
- (c) Make it turn left and right.
- (d) Turn its head left, right, up and down.
- (e) Make it watch your face (or a ball for the laser interface).
- (f) Enable the photo mode by touching its head.
- (g) Pronounce your first name.

- Mission

- (a) Draw the robot's attention toward one randomly chosen object among the 12 possible objects.
- (b) Take a picture of it.
- (c) Pronounce its predefined name.

The steps from (a) to (c) were repeated 4 times.

4. Fill in a post questionnaire.

The whole experiment (including the questionnaires) lasted from 20 to 30 minutes per participant.

### 4.4 Measures

During the experiments, we collected the pictures taken by the robot. They were automatically labeled. We also clocked of the time needed to complete the game and also the intermediate time, i.e. each time a picture was taken.

On top of these measures, we also conducted two questionnaire based surveys inspired by the classical guidelines in HRI. Before the experiment, we administered a demographic survey and a pre-task questionnaire concerning the participant's technological profile (computer, video games and robotic experience) and their attitude toward robotics. After the game, we conducted a post-task survey with the following assertions to which agreement had to be evaluated on a 5 points Likert scale:

- Usability and user's experience

- Q1 It was easy to learn how to use this interface.
- Q2 It was easy to move the robot.
- Q3 It was easy to make the robot look at an object.
- Q4 It was easy to interact with a robot.
- Q5 The robot was slow to react.
- Q6 Overall, it was pleasant to use this interface.

- Robotic game

- Q1 Completing the game was easy.
- Q2 The game was entertaining.
- Q3 I felt like cooperating with the robot.
- Q4 I picture myself playing other robotic games in the future.

## 5. RESULTS

We first manually sorted the collected pictures, i.e. training visual examples corresponding to a new object, into three categories: 1) images where the object was entirely seen; 2) images where the object was only partially present; 3) images where the object was not present at all. The objects were defined as "partially" as soon as a part was missing.

The figure 11 depicts these results. We performed a one-way ANOVA for the “entirely visible” condition and found a statistical difference between the four interfaces ( $F_{3,103} = 13.7$ ,  $p < 0.001$ ). In particular, we can notice that without providing any feedback about what the robot sees to the users (the Wiimote and gestures conditions), the object is entirely visible in only 50% of the images. The Tukey post-hoc test showed that providing a feedback significantly improves this result (80% for the laser and 85% for the iPhone). Furthermore, we can notice that the iPhone interface and in particular its video feedback allows users to almost never collect wrong learning examples (only 2%).

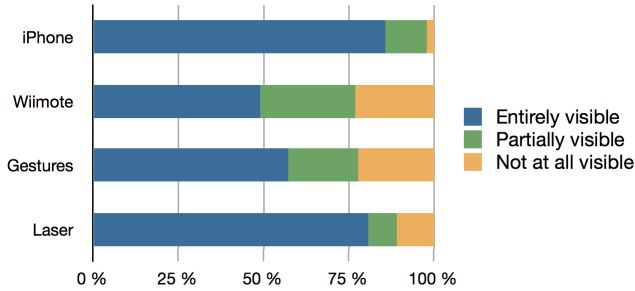


Figure 11: This figure shows the repartition of the collected images, among three categories: the object is 1) entirely, 2) partially or 3) not at all visible on the images. We can see that without any feedback (Wiimote or Gestures interfaces) the object were only entirely visible in only 50% examples. Providing a feedback significantly improves this result (80% for the laser and more than 85% for the iPhone).

The figure 12 presents the answers to the usability questionnaires. We performed one-way ANOVA and found statistical differences for the questions Q1 ( $F_{3,103} = 6.35$ ,  $p < 0.001$ ), Q2 ( $F_{3,103} = 2.44$ ,  $p < 0.05$ ), Q3 ( $F_{3,103} = 6.41$ ,  $p < 0.001$ ) and Q6 ( $F_{3,103} = 3.38$ ,  $p < 0.05$ ). The Tukey post-hoc tests showed that the iPhone, Wiimote and Laser interfaces were judged as easier to learn and more practical to move the robot than the Gestures interface. The users also stated that it was easier to make the robot look at an object with the iPhone and Wiimote interfaces. Furthermore, they also judged that overall the iPhone was significantly more pleasant to use than the Laser interface.

Figure 13 shows the results for the game part of the questionnaires. The only statistical difference was found for the question Q1. We can see that the participants found the game easier with the interfaces based on mediator objects than with the gestures interfaces ( $F_{3,103} = 5.17$ ,  $p < 0.005$ ). The game was judged as entertaining by participants for all conditions. It is also interesting to notice that the gestures condition seems to improve the feeling of cooperating with the robot. However, no statistical difference was found. Similarly, participants seemed to be more willing to play other robotic games in the future with the gestures condition than with the other conditions. These interesting results will be discussed in the next section.

As stated above, we also timed the experience. However, we did not find any significative difference among the different interfaces. Furthermore, for all the above results, no

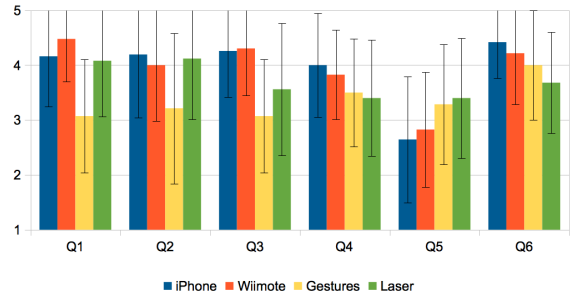


Figure 12: Usability: Participants found the gestures interface significantly less intuitive and harder to use than the other interfaces. They also stated that the iPhone interface was overall more pleasant than the Laser interface.

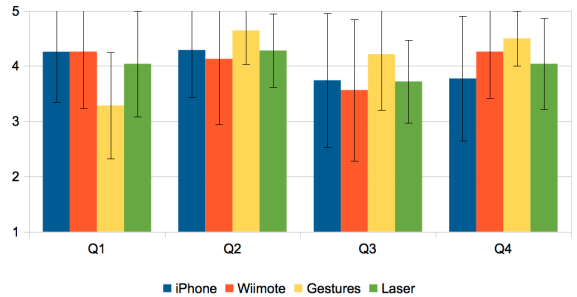


Figure 13: Robotic game: Our robotic game was stated as entertaining by all participants. They found the game significantly harder with the gestures interfaces but it increases the feeling of cooperating with the robot.

significant differences was found between the participants from the sciences museum and the participants from the university campus.

## 6. DISCUSSION AND CONCLUSION

In this paper, we showed the importance of the interface to allow users to robustly teach visual objects to a robot and in particular its impact on the quality of the learning examples they gathered. Indeed, we first showed that with simple interfaces such as the Wiimote or the Gestures **interfaces that do not provide any feedback to the users, users tend to collect only 50% of good learning examples**. This result seems particularly important to us as we are in a context where users would probably give very few learning examples. Thus, we can imagine a person giving four different examples of an object to its robot but only two will be actually usable by the recognition system. Then, we showed that specifically designed interfaces as the Laser and the iPhone interfaces, by notably **providing a feedback of what the robot perceives to the users, really significantly improve the quality of the learning example gathered**. While it was expected that providing a feedback to the users will help them to collect better learning examples, it is very interesting to notice that feedback such as the



visual feedback provided by the iPhone interface improve highly the quality of the learning examples and eliminate almost all the bad learning examples. Thus, if one is interested in allowing non-expert users to teach visual objects to a social robot by providing very few learning examples, we think that the interface should really be taken into consideration and specifically designed. In particular, as naive participants seem to have strong wrong assumptions about humanoids visual apparatus, we argue that the design of the interface should **not only help users to better understand what the robot perceive but should also drive them to pay attention to the learning examples** they are collecting. For instance, the iPhone interface presents on the screen of the device the learning example that users have to encircle. Thus, the interface naturally force them to monitor the quality of the examples they collected.

In this study, we also evaluated the user's experience of the robotic game. It is interesting to notice that while **the gestures interface was stated as less usable than the three other interfaces, participants judged that the game was as entertaining with this interface** as with the others. To us, this result can be explained by several factors. First, it is important to notice that the participants did not know whether they collected good learning examples or not. So, it did not influence their user's experience. For instance, users who collected only very bad learning examples could still think that they successfully finished the game. Second, while interfaces such as the iPhone interface were specifically designed to help users collect good learning examples, it was probably too complicated for a robotic game. Indeed, users had to monitor the robot, the game interface and the iPhone. Furthermore, it seems that the interface should be as transparent as possible in order to allow users to entirely focus on the game. Finally, the gestures interfaces seemed to improve the user's feeling that the robot is cooperating with them. We think that this result could be explained by the fact that participants were closer to the robot and that they were trying different gestures to see how the robot react and so try to determine which gestures were better understood. The bias introduced by the Wizard-of-Oz setup also lead to situation where the Wizard was adapting its behavior to the participants and thus was effectively cooperating with him. Although further studies should be carried out in this direction, our preliminary results seem to show that gestures interface could be interesting if one is interested in the development of simple robotic game.

## 7. REFERENCES

- [1] C. Breazeal and B. Scassellati. Infant-like social interactions between a robot and a human caregiver. *Adapt. Behav.*, 8(1):49–74, 2000.
- [2] A. G. Brooks, J. Gray, G. Hoffman, A. Lockerd, H. Lee, and C. Breazeal. Robot's play: interactive games with sociable machines. *Comput. Entertain.*, 2(3):10–10, 2004.
- [3] D. Filliat. Interactive learning of visual topological navigation. In *Proceedings of the 2008 IEEE International Conference on Intelligent Robots and Systems (IROS 2008)*, 2008.
- [4] K. Ishii, S. Zhao, M. Inami, T. Igarashi, and M. Imai. Designing laser gesture interface for robot control. In *Proceedings of the 12th IFIP Conference on Human-Computer Interaction, INTERACT2009*, pages 479–492, 2009.
- [5] F. Kaplan. *Les machines apprivoisées comprendre les robots de loisir*. vuibert, 2005.
- [6] F. Kaplan and V. Hafner. The challenges of joint attention. *Proceedings of the 4th International Workshop on Epigenetic Robotics*, 2004.
- [7] C. C. Kemp, C. D. Anderson, H. Nguyen, A. J. Trevor, and Z. Xu. A point-and-click interface for the real world: laser designation of objects for mobile manipulation. In *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 241–248, New York, NY, USA, 2008. ACM.
- [8] F. Lömker and G. Sagerer. A multimodal system for object learning. In *Proceedings of the 24th DAGM Symposium on Pattern Recognition*, pages 490–497, London, UK, 2002. Springer-Verlag.
- [9] D. R. Michael and S. L. Chen. *Serious Games: Games That Educate, Train, and Inform*. Muska & Lipman/Premier-Trade, 2005.
- [10] P. Rouanet, J. Béchu, and P.-Y. Oudeyer. A comparison of three interfaces using handheld devices to intuitively drive and show objects to a social robot : the impact of underlying metaphors. *RO-MAN*, 2009.
- [11] P. Rouanet, P. Oudeyer, and D. Filliat. Using mediator objects to easily and robustly teach visual objects to a robot. In *ACM SIGGRAPH 2010 Posters*. ACM, 2010.
- [12] 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.
- [13] S. S. Rumbaugh and R. Lewin. *Kanzi : The Ape at the Brink of the Human Mind*. Wiley, September 1996.
- [14] B. Scassellati. Mechanisms of shared attention for a humanoid robot. In *Embodied Cognition and Action: Papers from the 1996 AAAI Fall Symposium*, 1996.
- [15] L. Steels and F. Kaplan. Aibo's first words: The social learning of language and meaning. *Evolution of Communication*, 4(1):3–32, 2000.
- [16] M. L. Walters, S. N. Woods, K. L. Koay, and K. Dautenhahn. Practical and methodological challenges in designing and conducting human-robot interaction studies. In *Proceedings of the AISB'05 Symposium on Robot Companions Hard Problems and Open Challenges in Human-Robot Interaction*, pages 110–119. University of Hertfordshire, University of Hertfordshire, April 2005.
- [17] H. Wersing, S. Kirstein, M. Götting, H. Brandl, M. Dunn, I. Mikhailova, C. Goerick, J. J. Steil, H. Ritter, and E. Körner. A biologically motivated system for unconstrained online learning of visual objects. In *ICANN (2)*, pages 508–517, 2006.
- [18] M. Xin and E. Sharlin. Playing Games with Robots - A Method for Evaluating Human-Robot Interaction. *Education*, 2007.
- [19] H. A. Yanco, J. L. Drury, and J. Scholtz. Beyond usability evaluation: analysis of human-robot interaction at a major robotics competition. *Hum.-Comput. Interact.*, 19(1):117–149, 2004.