



HAL
open science

Toward a better understanding of the communication cues involved in a human-robot object transfer

Mamoun Gharbi, Pierre-Vincent Paubel, Aurélie Clodic, Ophélie Carreras,
Rachid Alami, Jean-Marie Cellier

► To cite this version:

Mamoun Gharbi, Pierre-Vincent Paubel, Aurélie Clodic, Ophélie Carreras, Rachid Alami, et al..
Toward a better understanding of the communication cues involved in a human-robot object transfer.
24th International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN,
Aug 2015, Kobe, Japan. pp.319-324, 10.1109/ROMAN.2015.7333626 . hal-01628249

HAL Id: hal-01628249

<https://hal.science/hal-01628249>

Submitted on 13 Dec 2018

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.

Toward a better understanding of the communication cues involved in a human-robot object transfer.

Mamoun Gharbi^{1,2}, Pierre-Vincent Paubel³, Aurélie Clodic^{1,4}, Ophélie Carreras³,
Rachid Alami^{1,4}, and Jean-Marie Cellier³

Abstract—Handing-over objects to humans (or taking objects from them) is a key capability for a service robot. Humans are efficient and natural while performing this action and the purpose of the studies on this topic is to bring human-robot handovers to an acceptable, efficient and natural level.

This paper deals with the cues that allow to make a handover look as natural as possible, and more precisely we focus on where the robot should look while performing it. In this context we propose a user study, involving 33 volunteers, who judged video sequences where they see either a human or a robot giving them an object. They were presented with different sequences where the agents (robot or human) have different gaze behaviours, and were asked to give their feeling about the sequence naturalness. In addition to this subjective measure, the volunteers were equipped with an eye tracker which enabled us to have more accurate objective measures.

I. INTRODUCTION

Fetch and carry abilities form the building blocks of what a service robot is intended to do. To support these abilities, a number of developments have been achieved in the human-robot interaction community, in terms handover location computation and task-oriented motion planning.

However, it appears that in front of a humanoid robot (with a head like a PR2), arm and base movement are not sufficient to achieve such tasks in a satisfying manner. We notice that people tend to anthropomorphize the robot and infer that the robot should use hand-eye coordination as humans do. This means that the robot should not only compute its arm movement but also a movement of its head. In this paper, we seek to find which head movement (or pattern of movements) should be added to robot arm movement to enable more natural interaction.

We focus on the way a robot should use its gaze cues (in our case head movement) to support a joint action involving robot arm movement. To achieve this goal, we conducted a multidisciplinary study involving psychologists and roboticists. We have chosen to study a simple task where a user (the giver) sets an object down on a table in the direction of another user (the receiver). Our study explores two cases for the giver: a human and a robot (Figure 1).

We have chosen to limit the study to a basic movement in order to concentrate on non-verbal cues that should be produced by the giver toward the receiver.

¹CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France
name.surname@laas.fr

²Univ de Toulouse, INSA, LAAS, F-31400 Toulouse, France

³CLLE-LTC UMR 5263 UT2J-CNRS, Toulouse, France
{pierre.paubel, carreras, cellier}@univ-tlse2.fr

⁴Univ de Toulouse, LAAS, F-31400 Toulouse, France



Fig. 1. Snapshots of the videos used to evaluate the movement naturalness.

A brief analysis of this kind of gesture between two humans shows the significance of two variables to judge this feature:

- Speed of the gesture (more or less fast movements) and
- Gaze direction of the giver toward the receiver

Our task is not a handover but consists in setting down the object, and our aim is to find which “coordination smoothers” ([1], [2]) should be added to the robot motion to look more natural to the human it interacts with. Use of gaze cues have already been studied in HRI context [3], [4], [5], [6], [7] and this paper should be considered as complementary to these studies. We seek to confirm, in a controlled experimental context, results that have been already exhibited. In addition, what could be credited to our work is the study of gaze behaviour for both the giver (H or R) and the receiver during a give action. The closest studies in the field to our knowledge are [5] and [7]. In the first one, authors do a similar study over a different action, while the second does not analyze the receiver gaze.

II. GENERAL HYPOTHESIS

Our first general hypothesis comes from our subjective experience: the movement will be considered as more natural if it includes the use of gaze cues, e.g. a robot with a fixed head will be considered less natural than one with a moving head. We have also considered the sequence of the gaze focus steps (the order in which object and receiver are looked at)

to see if a specific pattern emerges, with a human or a robot giver.

Our second hypothesis is that gaze movement of the receiver will be directed toward the head and toward the object. We want to see here if the head and the object of a man or a robot giver are looked at in the same way.

Our third hypothesis is exploratory and concerns the effect of the movement speed with respect to the evaluation of its naturalness.

III. RELATED WORKS

There have been a number of contributions concerning motion generation for a humanoid robot and more specifically for its arm(s) in a human-robot interaction context.

A. Human-Aware Motion Generation

Sisbot et al. [8] presented a general motion generation framework for a manipulator robot to produce a robot behaviour that can be called social in the sense that it tends to respect social rules. This framework has been validated by a user study (Dehais et al. [9]) and has been extended in Mainprice et al. [10] by taking into account, at handover motion planning level, how the human can participate to the task.

Cakmak et al. [11] focused on the choice of handover configuration by an evaluation of how humans would prefer being handed an object by a robot. They found that configurations learned from the human were preferred over planned configurations (even if these latter provide better reachability of the object). They concluded in saying that while a planning approach has the potential to produce configurations that are practical, it was insufficient in addressing usability, naturalness and appropriateness for the task.

Strabala et al. [4] attempted to codify a procedure for seamless human-robot handovers through a handover structure. The authors proposed that the robot should be able to handle social norms and should be readable enough to enable the human to infer the what, when and where of the handovers (as it is proposed in a more general way for joint action by Sebanz and Knoblich [12]). They proposed e.g. a turn-taking pattern based on giver actions, face direction and gaze. They found that people could easily understand human-like cues performed by a robot, and that they preferred these cues to machine-like ones. Among these cues, the gaze is an important one.

B. Gaze Management

Gaze analysis allows the receiver to make hypothesis on the cognitive activity handled by the giver, and a number of researchers tried to codify and implement this cues on robots.

Mutlu [3] studied the use of social gaze cues in communication on several robotics platforms (ASIMO, Robovie and Geminoid) and showed the importance of gaze in such a context and that the use of well-defined gaze patterns could enhance human-robot communication experience.

Boucher et al [5] observed that one of the current road-blocks in the elaboration of smooth and natural human-robot

cooperation is the coordination of robot gaze with the ongoing interaction and tried to identify pertinent gaze cues in human-robot cooperation. They found that humans could reliably exploit robot gaze to allow them to perform in an anticipatory manner in a cooperative task.

Interestingly, in a study oriented toward gaze cues in human-human interaction, Furlanetto et al. [13] showed that eliminating gaze cues by blurring the actors face did not reduce perspective-taking, suggesting that in the absence of gaze information, observers rely entirely on the action. Intriguingly, perspective-taking was higher when gaze and action did not signal the same intention, suggesting that in presence of ambiguous behavioural intention, people are more likely to take the others perspective to try to understand the action.

Staudte et al. [6] explore visual attention in spoken human-robot interaction and found that robot gaze which was congruent with an uttered sentence helped human interlocutors to faster judge utterances than if robot gaze was absent (and on the other hand, when robot gaze was incongruent with the utterance, it slowed people down).

Moon et al. [7] exploited human-like gaze cues during human-robot handovers and found on one side that subjects reached for the object significantly faster when the robot directed its gaze toward the intended handover location than when no gaze cues were used. On the other side, subjects tended to perceive handover as more natural when the robot provided turn-taking gaze.

This shows the importance of gaze in such task. However what this could mean? In fact, the robot would be able to achieve the task without gaze; it has only to control its arm. But this seems not enough.

C. About Joint Action

Let us try to situate this work in the global frame of joint action. Sebanz et al. [14] proposed as a working definition: “joint action can be regarded as any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment”. They propose that successful joint action depends on the abilities “(i) to share representations, (ii) to predict actions, and (iii) to integrate predicted effects of own and others actions”.

Vesper et al. [1] established that a minimal architecture for joint action should be able to handle, besides the goal, tasks representation (possibly shared), monitoring and prediction processes, what they call coordination smoothers. They argued that “where joint action requires precise coordination in time or space, there are often limits on how well X’s actions can be predicted. One way to facilitate coordination is for an agent to modify her own behaviour in such a way as to make it easier for others to predict upcoming actions.”

In the same direction, Becchio et al [2] showed that it is possible to differentiate the kinematics of an action performed by an agent acting in isolation from the kinematics of the very same action performed within a social context. They explain that social context shapes action planning and

that, in the context of a social interaction, flexible online adjustments take place between partners.

We suggest that gaze cues could hold this role of coordination smoother in helping the human in front of the robot to better understand robot behaviour and help the robot to achieve its movement in a more natural way.

IV. MATERIALS AND METHODS

A. Participants

33 volunteers participated to the experiment (age range 22–38, $M=27$, $SD = 3.5$; 21 males, 12 females). There were not familiar to robots or to work with a robot. 15 watched human videos and 18 watched robot videos. All participants had normal uncorrected vision. Two volunteers have to be excluded from subsequent analyses due to a technical problem that damaged eye-tracking data (unreliable calibration).

B. Experimental Set-up

The experimental situation implies watching a video where a giver (Human or Robot), seated behind a table, takes the object with his right hand, and put it on the table so that the receiver, behind the video camera, can reach it. We use an experimental setup quite similar as the one of Staudte et al [6]. This setup can be found not interactive enough (we will come back to this choice in section VII). In this study, we are trying to isolate gaze cues and movement velocity in order to find some hints about the use of these factors in our context. Moreover, it has been proposed by Kiesler [15] and Woods [16] that video-based scenario can enable us to infer valuable results.

The experiment took place in a room where temperature and luminosity were kept constant. Participants faced a computer screen where the video was presented. Eye movements were recorded using an EyeLink 1000 remote eye tracker (SR Research Ltd., Mississauga, Ontario, Canada). This eye tracker possesses a spatial accuracy greater than 0.5° and a 0.01° spatial resolution. The sampling rate was set to 1000 Hz. The camera was placed at a distance of 20 cm from the screen and the eye-camera distance was 60 cm. We used a display screen DELL 19" with a refresh rate of 75 Hz and a resolution of 1024x768 pixels. A forehead rest was used to maintain these distances and to avoid heads movements. All eye tracking data were extracted using the SR Research default centroid algorithm. Simulation room enlightenment was maintained constant (19 lux).

In the experiment, we manipulated 3 variables: (1) the type of giver (Human or Robot), (2) the speed of the movement and (3) the gaze behaviour. The Robot was a PR2 and the Human was a white man (65 years old). Videos were shot to be as similar as possible (see figure 1). Three different speeds of the arm movement were presented: normal, rapid, and slow. Videos were accelerated and decelerated so that the movement remains relatively natural. The duration of the slow (3750 ms), normal (2250 ms), and rapid videos (1750 ms) were identical for the human and for the robot.

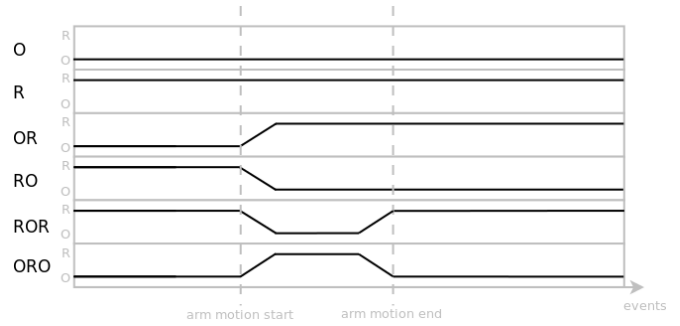


Fig. 2. the timing of the different gaze behaviours according to were the givers looks.

Six different sequences were created for the gaze behaviour, their timelines are displayed figure 2:

- **O** the giver looks only at the object, while the object is moving, the giver looks at the center of the object
- **R** the giver looks only at the receiver
- **RO** the giver looks at the receiver at the beginning, then when its arm (and the object in it) starts moving it starts to look at the object and stays on it till the end
- **ORO** the giver looks at the object at the beginning, then when its arm is moving it starts to look at the receiver till the end of the motion. When the motion is over it looks at the object again till the end
- **OR** the giver looks at the object at the beginning, then when its arm is moving it starts to look at the receiver and keeps this position till the end
- **ROR** the giver looks at the receiver at the beginning, then, when its arm is moving (and the object in it) it starts to look at the object till the end of the motion. When it is over it looks at the receiver again till the end.

The human moved only his eyes whereas the robot (PR2) moved its head to simulate a gaze. Imai and al. [17] established that the perception of the robot gaze is coupled to the robots head orientation. Examples of those videos are available here <http://homepages.laas.fr/magharbi/userstudy.html>

C. Procedure

Participants were told that they had to evaluate a movement naturalness presented on a video. We made the choice, for methodological reasons, to use a very neutral formulation to not influence the judgment of the user. The same choice have been done concerning the question asked. It is one of the three that has been used in [7], whereas in [5] there were no subjective evaluation.

Participants were familiarized with the situation by watching a first video, a bit different from the experimental ones. The experimental session was composed of 18 trials: 3 (speed) x 6 (gaze). In a trial, the participant pressed a button to begin the video. Immediately after the video, he/she rated the perceived naturalness of the movement on a 5 points Likert scale presented on the screen. Between trials, participants had to complete a digital logical suite, to break the dullness of the task. The order of trials was

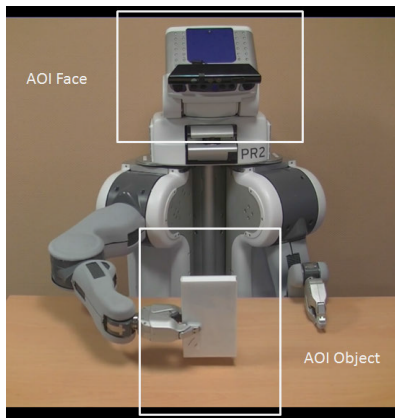


Fig. 3. The AOIs used in the oculometric measurement

randomized. The session (18 trials) was repeated one time. Thus, participants watched and rated 36 videos in all.

D. Subjective measurement

To assess participants subjective experience, they had to rate perceived naturalness after each video, with a 5-point scale (5 for perfectly natural, 1 for not natural at all).

E. Oculometric measurement

Classical dependent variables in eye-tracking studies include the number and duration of fixations on areas of interest. In this study, the areas of interest (AOIs) were (1) the giver's face and (2) the object. Those AOIs were static, as depicted in figure 3. As video duration changed between experimental conditions, we computed the percentage of dwell time spent on AOIs to study the distribution of the visual attention.

F. Instructions

We considered that viewing the videos, the interpretation of the gesture objective was clear so no comments were done on that aspect to the participants. The instruction and the question were very neutral, in order to not influence the user judgement. They were only asked to rate the previous sequence naturalness as explained section IV-D. Note that the instruction and questions were written in the participant mother tongue (French)

V. RESULTS

We performed a mixed-design analysis of variance to examine the effects of (1) the gaze behaviour, (2) the speed movement, and (3) the type of giver on our subjective and oculometric dependent variables. Subjective and eye tracking data have been analysed with the software package Statistica 8.0 (Statsoft, Tulsa, Ok, USA).

A. Subjective measurements

Gaze Behaviour:

Results indicated a main effect of the gaze behaviour on the naturalness ratings, $F(5, 145)=15.034$, $p<.001$ (figure 4). Post-hoc paired comparisons showed that OR and ROR gaze behaviour are significantly judged more natural than the

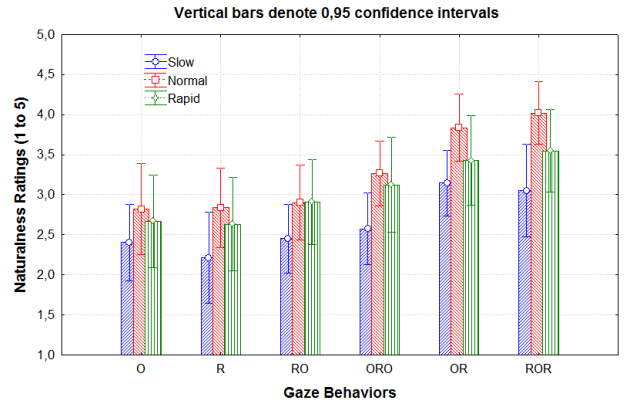


Fig. 4. Naturalness ratings as a function of the gaze behaviour and the speed movement

four others conditions R, O, RO, ORO (highest p-value in the post hoc table equal to .003). No difference was found between (1) the two conditions OR and ROR ($p=.70$) and (2) the three conditions R, O and RO (lowest p-value equal to 0.48). Finally, the condition ORO is significantly judged more natural than the two gazes behaviour R and O (highest p-value equal to .03).

Movement speed:

Results indicated a main effect of the movement speed on the naturalness ratings, $F(2, 58)=10.354$, $p<.001$. Fishers LSD post-hoc comparisons showed that the normal and the rapid speed movement conditions are judged more natural than the slow speed movement condition (highest p-value in the post hoc table equal to .004). No significant difference was found between the normal and the fast speed conditions ($p=.16$).

Type of giver (Human vs. Robot):

Interestingly, there was no significant difference in the results between the two types of givers ($F(1, 29)=1.988$, $p=.16$). Moreover, no interaction was found between the three main manipulated factors. This result suggests that the effects of (1) movement speed and (2) gaze behaviour described above are not influenced by the type of giver (human or robot).

B. Eye tracking measurements

Distribution of the visual attention between the face of the giver and the object:

Overall, results indicated a significant difference between the mean percentage of dwell time spent on the face of the giver and the mean percentage of dwell time spent on the object ($F(1, 29)=59.848$, $p<.001$). Participants tend to focus mainly their visual attention on the face of the giver.

The detailed analysis (see figure 5) showed an interaction between the type of giver and the gaze behaviour. Firstly, when the giver is a human, there is no effect of the giver gaze behaviour on the mean percentage of dwell time spent on the face of the giver ($F(5, 65)=0.807$, $p=.54$), nor on the mean percentage of dwell time spent on the object ($F(5, 65)=1.004$, $p=.42$). The main result here is that, when the giver is human, participants focus mainly their visual attention on the face of the giver to provide a judgement concerning the naturalness of the task, independently of the

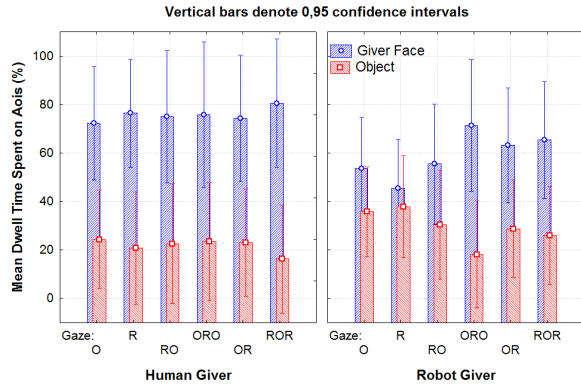


Fig. 5. Distribution of the visual attention between AOIs as a function of the gaze behaviour and the type of giver

giver gaze behaviour. Secondly, when the giver is a Robot, results indicated a main effect of the gaze behaviour on the mean percentage of dwell time spent on the face of the giver ($F(5, 80)=12.82, p=.001$), and on the mean percentage of dwell time spent on the object ($F(5, 80)=6.264, p=.001$). Fishers LSD posthoc comparisons showed that participants focus more on the face of the robot for the three types of gaze behaviour ORO, OR and ROR than for the three others conditions R, O, RO (highest p-value in the post hoc table equal to .04). On the other hand, participants focus less on the object for the same three types of gaze behaviour ORO, OR and ROR than for the two others conditions R, O (highest p-value in the post hoc table equal to .05). However, gaze behaviours OR and ROR are not significantly different of the condition RO (lowest p-value equal to 0.26). If we take the human giver as a reference, the oculometric pattern with a robot giver is identical to the one of a human giver only for the ORO, OR and ROR conditions. Finally, no effect of the movement speed was found ($F(2, 58)=1.798, p=.43$), and the reader may find useful to know that there was no difference between the two blocks of video presentation ($F(1, 29)=0.947, p=.33$).

Pupil size:

Results indicated a strong main effect of the type of giver on pupil size variations, with larger pupil diameters when the giver is a robot ($F(1, 29)=12.803, p<.001$). Results also revealed a main effect of the gaze behaviour ($F(5, 145)=3.050, p<.001$). However, post-hoc paired comparisons showed only one significant difference, with smaller pupil size in the OR gaze behaviour condition (highest p-value equal to .03 in the post hoc table). No effect of the movement speed was found ($F(2, 58)=1.798, p=.17$). Finally, results indicated a significant difference between the two blocks of video presentation ($F(1, 29)=26.155, p<.001$), with smaller pupil size during the last block, what could be reasonably considered as a training effect.

VI. DISCUSSION

The results presented in section V give a number of hints about humans preference in our setup.

A. First Hypothesis

Subjective measurement concerning gaze behaviour confirms that the behaviour seems more natural when the giver uses its gaze, more precisely **OR** (the giver looks at the **Object** then at the **Receiver**) and **ROR** (the giver looks at the **Receiver** then at the **Object** and then at the **Receiver** again). This first result strengthens our first hypothesis: when the giver does not send gaze cues (**O** case (the giver looks only at the **Object**), and **R** case (the giver looks only at the **Receiver**), the interaction lacks naturalness. However, gaze movements order seems to affect as well these results: **ROR** and **OR** are preferred to **ORO** (the giver looks at the **Object** then at the **Receiver** then at the **Object** again), which is preferred to **RO**. We observe a **OR** (**Object** then **Receiver**) pattern at the end of the movement that could be considered as an acknowledgement/turn-taking signal from the giver to the receiver. Note that the variable “type of giver (human or robot)” does not affect the subjective naturalness rating.

B. Second Hypothesis

Concerning the distribution of visual attention between the giver’s face and the object, results are different according to the type of giver. When the giver is a human, there is no effect of the giver gaze behaviour on the distribution of the visual attention. The receiver focuses mainly his visual attention on the face of the human. We believe this behaviour is normal for humans as the face is the most expressive part of the body and humans are used to focus on the face to determine a number of features. When the giver is a robot, we can distinguish two cases:

O, R and RO cases: The receiver visual attention is shared between the face and the object. That means, receiver will not focus either on the face or on the object but may go from one to the other. We interpret this as the receiver being lost in this kind of situations. Further analysis of eye tracking data is needed to validate this interpretation.

ORO, OR and ROR cases: The visual attention is mainly focused on the head of the robot. In those cases, we found the same pattern of visual attention as in the human giver.

Consequently, our second hypothesis is not validated (it holds only for movements judged as less natural, otherwise receiver visual attention is mostly focused on the giver).

C. Third Hypothesis

Concerning the speed of movement, the slow one is considered less natural than the rapid and normal ones. This study was not meant to study deeply this aspect. However, this corroborates our third hypothesis which is in fact an intuitive observation of roboticists that robot slowness is an issue when we have to evaluate it.

D. General results

Taken together, the results on the perceived naturalness of the movement and the ones on the oculometric pattern of the receiver seem to put forward two main conditions: **OR** and **ROR**. Those two conditions are not only perceived as more natural than the others (with a robot or a human

giver) but they present a similar oculometric pattern of the receiver (with a human or a robot giver). It seems that the final **OR** is important. When the giver, at the end of the movement, moves the gaze from the object to the receiver, it may signal the end of the exchange. The fact that the receiver looks mainly at the face also in the **ORO** condition may be interpreted in the same sense: when the robot ends its movement on the object, the receiver seeks an acknowledgement on the robot head (our first look at more detailed eye-tracking results seems to corroborate this interpretation).

Our study is more about the movement itself rather than its initiation, however, our preferred patterns meet the ones found in [4]. That is at the beginning of the action, the robot is looking at the object or at the receiver. We also showed that the gaze at the end of the exchange seems also important. We found that some patterns are considered as more natural than others, whereas [7] did not find any difference on that aspect. In the robot case, this is confirmed by objective measurements. These patterns tend to confirm our first intuition and findings about handover condition [7], [4] or in a cooperative task [5] that stated that a human exploits the gaze of the robot when it is present.

The difference in the pupil size between the two types of givers (human or robot) might have different explanations: more curiosity or cognitive load induced by the observation of non-familiar, unknown machine. In the general eye-tracking literature, pupil diameters have been found to increase along with cognitive demands [18] and emotional load [19]. In this context, the difference in the pupil size between the two types of givers (human or robot) might have different explanations: more curiosity or cognitive load induced by the observation of non-familiar, unknown machine.

VII. CONCLUSION AND FUTURE WORK

In this paper, we seek to find which head motion or pattern of head motions should be added to arm motion to enable more natural interaction. This study is a joint initiative between a robotics and a psychology team. We have made the deliberate choice to use videos instead of direct interaction (H-H or H-R) following methodological considerations. It is difficult to reproduce exactly the same behaviour for several experimental subjects of a given condition and in the mean time enough heterogeneous and stable between experimental conditions. Through this choice, we give priority to methodological quality of experimental conditions comparisons to the detriment of the ecological validity. However, we have made this choice because we believe that such a study, with controlled experimental conditions, can help the HRI community to build upon a common basis and help check which questions have been already answered (and with which answers) and which questions are still open.

We have found that **OR** (the giver looks at the **Object** then at the **Receiver**) and **ROR** (the giver looks at the **Receiver** then at the **Object** and then at the **Receiver** again) are preferred patterns for our task. We plan to integrate such results into robot motion planning algorithm and to test it

on a real robot. We will also carry on the analysis of eye-tracking results to study more precisely where the receiver is looking at, and when. This will serve two goals. The first one is to analyse if this could help us to draw shared attention gaze pattern during the movement. The second one is to get data about where and when a receiver looks at in order to integrate such knowledge onto a receiver robot.

ACKNOWLEDGMENT

This work was conducted within the EU SAPHARI project funded by the E.C. division FP7-IST under contract ICT-287513, and within “Coopération et Interaction Robot-Humain” (CIRH), Maison des Sciences de l’Homme et de la Société de Toulouse, (MSHST), Université Fédérale de Toulouse Midi-Pyrénées.

REFERENCES

- [1] C. Vesper, S. Butterfill, G. Knoblich, and N. Sebanz, “A minimal architecture for joint action,” *Neural Networks*, 2010.
- [2] C. Becchio, L. Sartori, and U. Castiello, “Toward you the social side of actions,” *Current Directions in Psychological Science*, 2010.
- [3] B. Mutlu, “Designing gaze behavior for humanlike robots,” Ph.D. dissertation, Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, 2009.
- [4] K. W. Strabala, M. K. Lee, A. D. Dragan, J. L. Forlizzi, S. Srinivasa, M. Cakmak, and V. Micelli, “Towards seamless human-robot handovers,” *Journal of Human-Robot Interaction*, 2013.
- [5] J.-D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, P. F. Dominey, and J. Ventre-Dominey, “I reach faster when i see you look: gaze effects in human-human and human-robot face-to-face cooperation,” *Frontiers in neurobotics*, 2012.
- [6] M. Staudte and M. W. Crocker, “Visual attention in spoken human-robot interaction,” in *ACM/IEEE, Human robot interaction, HRI*, 2009.
- [7] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zeng, B. A. Blumer, K. MacLean, and E. A. Croft, “Meet me where i’m gazing: how shared attention gaze affects human-robot handover timing,” in *ACM/IEEE, Human-robot interaction, HRI*, 2014.
- [8] E. A. Sisbot and R. Alami, “A human-aware manipulation planner,” *Robotics, IEEE Transactions on*, 2012.
- [9] F. Dehais, E. A. Sisbot, R. Alami, and M. Causse, “Physiological and subjective evaluation of a human-robot object hand-over task,” *Applied ergonomics*, 2011.
- [10] J. Mainprice, M. Gharbi, T. Siméon, and R. Alami, “Sharing effort in planning human-robot handover tasks,” in *IEEE, Robot and Human Interactive Communication, RO-MAN*, 2012.
- [11] M. Cakmak, S. S. Srinivasa, M. K. Lee, J. Forlizzi, and S. Kiesler, “Human preferences for robot-human hand-over configurations,” in *IEEE/RSJ, Intelligent Robots and Systems, IROS*, 2011.
- [12] N. Sebanz and G. Knoblich, “Prediction in joint action: What, when, and where,” *Topics in Cognitive Science*, 2009.
- [13] T. Furlanetto, A. Cavallo, V. Manera, B. Tversky, and C. Becchio, “Through your eyes: incongruence of gaze and action increases spontaneous perspective taking,” *Frontiers in human neuroscience*, 2013.
- [14] N. Sebanz, H. Bekkering, and G. Knoblich, “Joint action: bodies and minds moving together,” *Trends in cognitive sciences*, 2006.
- [15] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey, “Anthropomorphic interactions with a robot and robot-like agent,” *Social Cognition*, 2008.
- [16] S. Woods, M. Walters, K. L. Koay, and K. Dautenhahn, “Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach,” in *IEEE, Workshop in Advanced Motion Control*, 2006.
- [17] M. Imai, T. Kanda, T. Ono, H. Ishiguro, and K. Mase, “Robot mediated round table: Analysis of the effect of robot’s gaze,” in *IEEE, Robot and Human Interactive Communication, RO-MAN*, 2002.
- [18] D. Kahneman and J. Beatty, “Pupil diameter and load on memory,” *Science*, 1966.
- [19] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, “The pupil as a measure of emotional arousal and autonomic activation,” *Psychophysiology*, 2008.