Robot initiative increases the rhythm of interaction in a team learning task

[Extended Abstract] *

Serena Ivaldi †
Institut des Systèmes Intelligents et de Robotique (ISIR), Paris, France
serena.ivaldi@isir.upmc.fr

Salvatore M. Anzalone
Institut des Systèmes Intelligents et de Robotique (ISIR), Paris, France
anzalone@isir.upmc.fr

Woody Rousseau
Institut des Systèmes Intelligents et de Robotique (ISIR), Paris, France
woody.rousseau@ensta-paristech.fr

Olivier Sigaud
Institut des Systèmes Intelligents et de Robotique (ISIR), Paris, France
sigaud@isir.upmc.fr

Mohamed Chetouani
Institut des Systèmes Intelligents et de Robotique (ISIR), Paris, France
chetouani@isir.upmc.fr

ABSTRACT

We hypothesize that the initiative of a robot during a collaborative task with a human can influence the pace of interaction and the reaction time of the human response to attention cues. We designed a two-phases object learning experiment where the human teaches the robot about the properties of some objects. We compare the effect of the initiator of the task in the teaching phase (human or robot) on the rhythm of the interaction in the verification phase. We measure the reaction time of the human gaze when responding to attention utterances of the robot. Our experiments show that when the robot is the initiator of the learning task, the pace of interaction is higher and the reaction to attention cues faster.

*A full version of this paper is available as *Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. Frontiers in Neurorobotics. Vol 8, No 5, DOI 10.3389/fnbot.2014.00005.* The work was partially funded by the European Commission, through the Projects CoDyCo (FP7-ICT-2011-9, No. 600716) and Michelangelo (FP7-ICT No. 288241), and by French National Agency of Research, through the Projects MACSi (ANR-10-BLAN-0216) and EDHHI/SMART (ANR-11-IDEX-0004-02).

†All the authors are with:
Sorbonne Universités, UPMC Univ Paris 06, UMR 7222, Institut des Systèmes Intelligents et de Robotique (ISIR), F-75005, Paris, France
CNRS, UMR 7222, Institut des Systèmes Intelligents et de Robotique (ISIR), F-75005, Paris, France

1. INTRODUCTION

Verbal and non-verbal behaviors, interplays and interpersonal synchrony are critical to determine the functional role of humans during social tasks [1]. Non-verbal communication is critical for achieving natural interaction between the human and the robot, especially to achieve engagement and synchronization in tasks where the two are teamed. Implicit non-verbal communication positively impacts human-robot tasks performance [2]. The inter-personal synchrony and the timing of the attention response during HRI can be used as an indirect measure of the human engagement toward the robot and the task [3, 4]. Particularly in turn-taking scenarios where the partners collaborate in a systematic way, time is critical [5]: synchrony, delays in replies, and rhythm of interaction can impact on the perception of the robot by the human, amplifying or decreasing the perceived engagement and influencing impression and responsive behaviors. This is true for robots [6] and agents in general [7]. Therefore, there is an apparent link between the timing of interaction, the role of the partners in the interaction and their impressions and responsive behaviors: while this has been shown for virtual conversational agents [8], there are still open questions about robots in teaching, collaborative and conversational scenarios. Here, we hypothesize that the reaction time of a human in response to a robot utterance can depend on the roles of the human and robotic partners during the interaction. We also hypothesize that the different roles of the partners can influence the rhythm of interaction and the perceived engagement [9]. We study the effects of a simple joint attention system during a learning task, in terms of induced joint attention (reaction time) and engagement perceived on the human side. We designed a two-step learning experiment, where people had to teach the iCub the color of objects. The learning situation was simple, as shown in Fig. 1, and interaction between robot and human was naturalistic. Remarkably, participants were not required to do any calibration nor wear eye-tracking devices, such as in [10]. Objects were selected by simply gazing at them by both pairs, and the information transfer about their color property was based on verbal communication (speech). We compare the effect of the initiator of the task in the teaching phase (human or robot) on the reaction time of the human gaze when responding to attention cues of the robot and the rhythm of interaction.

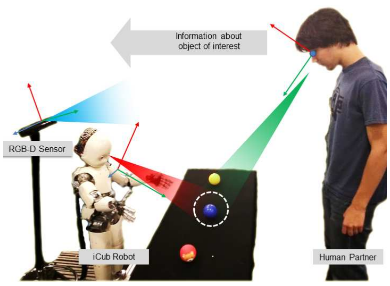

Figure 1: The object learning experimental scenario.

2. MATERIALS & METHODS

Experimental protocol. The experimental scenario is shown in Fig. 1: the human is standing in front of the robot, between the two there is a table with several colored objects. The experiment consists of a supervised learning process, across two phases: a teaching phase and a verification phase. In the first phase, the robot is taught the labels of the objects by the human partner. The object of interest is selected through joint gaze. In the second phase, the human gazes to one of the objects, and the robot responds with the learnt label. The teaching phase can be performed in two different conditions: *Human Initiative* (HI) and *Robot Initiative* (RI). HI and RI conditions are used to establish which partner initiates the action, that is the first who gazes at the object of interest.

Experimental procedure. We recruited 13 adult volunteers within the local campus population, mostly from the ISIR laboratory, who had no prior experience of interactions with iCub. Volunteers were divided into two groups, and associated to condition HI or RI. The only requirement for the participants was the lack of prior experience with the robot; that is we selected volunteers that had never interacted with the robot before. Age and sex were not control variables for our study, and interpersonal differences were not relevant. Participants were equally informed about the purpose of the study and the technological limitations of the robot. They were equipped with a lavalier microphone, and then could enter the robot's room. They were simply instructed to stay in front of the robot (their position was not fixed a priori) and do the teaching task. No calibration was required. iCub was not moving except for the head, which was a choice made to avoid proactive gaze effects. Three colored balls were placed on a table in between the human and the robot, on the left, on the center and on the right. During the two phases, participants were free to speak and interact with the robot in the way they were feeling more comfortable with. The teaching session was composed in both HI and RI conditions of only 3 trials, corresponding to the 3 objects to teach. Once the 3 colors were learnt, the verification phase would begin: the humans needed to verify that the robot had learnt correctly the 3 objects.

Robotic framework. The robot behavior was controlled by a pool of software modules developed in YARP and ROS [11], particularly for robot gaze, 3D people tracking, head pose estimation, object recognition, verbal communication. Technical details about these modules can be found in [12].

Measurements. The course of the experiment was controlled by a finite state machine. The timing of the events generated by the computer, the robot actions and the participants' responses were imported from the log of each experiment. The human gaze was continuously estimated from the RGB-D sensor through a gaze tracking module. Overall, gaze strategies for the HI and RI groups were diverse, but while the timing of the reactions was critical, the inter-individual differences about gaze were not relevant for our study. We retrieved two important measurements. The first is the reaction time of the human in response to the attention stimulus of the robot, i.e., the request to select an object. We measure in this case the time elapsed between the onset of robot speech and the time when the human gaze, stabilized on the object of interest, is correctly identified by the robot. The second is the interval between two successive requests from the robot, marking the amount of time dedicated by the partners to exchange information about the object of interest. This measurement is inversely proportional to the *pace of interaction* as it has been defined by [4]. The shorter the interval, the higher the pace or the faster the rhythm.

3. RESULTS

Table 1 and 2 report the reaction time and the indirect pace measurement for the participants of the two groups. The time distributions were compared with Wilcoxon's test. The test shows that there is a difference in the timing between the two groups ($p \leq 0.005$). **People in the RI group react faster than the ones of the HI group, and the interaction with the robot has a higher rhythm** (see Figure 2). Figure 3 shows the normalized gaze heat-maps of the two groups. Each map is a plot in the head's pitch-yaw space, thus each point represents the gaze direction of the human during the interaction with the robot. The range of pitch and yaw is $[-90, 90]$ degrees. For the head pitch, 90 is on top of the head, 0 is in front of the head, -90 is below the head. For the yaw, 0 is in front of the head, while -90 and 90 represent left and right. We identified the four clusters associated to the robot head and the three objects by applying K-means on the points, indicated in the left upper corner of each plot. We compared the density of each cluster in the two conditions with Wilcoxon's test. The test showed that there is no significant statistical difference in the clusters for both conditions ($p > 0.1$). It is however interesting to observe the amount of time spent by the participants in looking at the different salient topics, which is proportional to the density of the clusters. Overall **humans spent** 66% **of their time looking at the robot**. For the three objects, the amount of time is unequal: while the left and right objects get almost the same amount of time (7% and 6%), the object in the center was the focus of attention for almost twice the time spent for the others (21%). This has a double explanation: on one side, it is more difficult for the robot to detect that the human has moved the head to gaze at the object of interest if the movement is exclusively on the head pitch; on the other side, sometimes participants spontaneously looked downward to match the robot's gaze (this behavior is in fact "normal" for humans and rather a

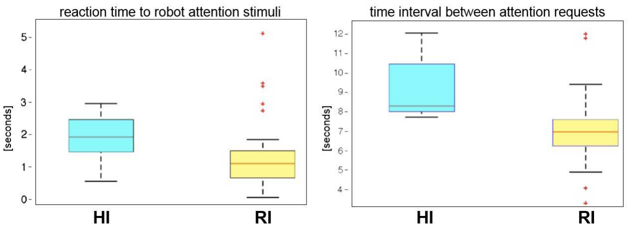

Figure 2: Reaction time to robot attention stimuli and time interval between consecutive attention requests in the verification phase.

Table 1: Reaction time (seconds) in response to robot attention stimuli (utterances) during verification phase

Group	mean	std	median	Wilcoxon's test
HI	1.932	0.711	1.917	W=418,
RI	1.296	1.145	1.106	p-value=0.005

Table 2: Time interval (seconds) between consecutive robot attention stimuli (utterances) during verification phase

Group	mean	std	median	Wilcoxon's test
HI	9.524	1.515	8.588	W=447;
RI	7.287	1.653	7.257	p-value=1.6e-5

positive sign of natural, engaged interaction with the robot).

4. DISCUSSION

The response times reported in Table 1 show that humans respond faster to robot's utterances in the verification phase when in the previous phase of the task the robot was leading the interaction (RI condition). The measurements verify our initial hypothesis, that is the difference in the initiator/leader of the learning task in the first phase is reflected in different reaction times in the second phase of the task. In the HI teaching phase, the robot asks the human to choose an object, leaving the choice to the human, and making him the main actor of the interaction. Once the human has gazed to the object, and its gaze is correctly estimated, the robot looks at the object of interest. The rhythm of the interaction is essentially determined by the human response to the robot's utterance: in terms of time, the human can move more or less quickly his head, and make the movement more or less "readable" by the robot, thus influencing the time needed by the robot to estimate the head direction correctly. Once the direction is estimated, the robot moves its eyes and head with a practically constant movement, determined by the gaze controller - the same used in [13]. In the RI teaching phase, the robot randomly picks an object on the table and asks the human to tell the color of the object. The choice in this case is made by the robot, which initiates the interaction. The rhythm of the interaction as well as its success is determined by the readability of the robot, its capability to induce in the human a prompt response to the robot attention request, and of course by the readability of the human that needs to have the same referential focus as the robot to

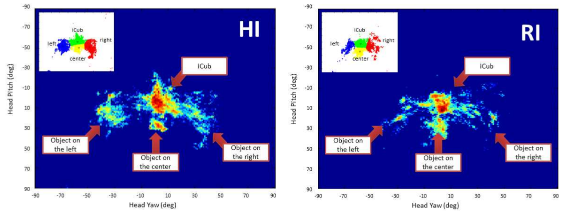

Figure 3: Normalized gaze heat map of the human partners in the HI and RI groups. The plots show the points in the pitch-yaw space representing the gaze direction of the human partners during the interaction with the robot. Note that the range of pitch and yaw is $[-90, 90] degrees$. For the head pitch, -90 is on top of the head, 0 is in front of the head, 90 is below the head. For the yaw, 0 is in front of the head, while -90 and 90 represent left and right.

make the interaction advance[1]. Again, the duration of the robot's movements is fixed, so the human is the main actor responsible for setting the pace of the interaction through his behavior. Why do these two conditions reflect in different reaction times in the verification phase? There could be several reasons. One possible reason is that in the RI condition, participants learned how to "read" the robot behavior to advance in the teaching phase, and reply to its questions contingently. Therefore, in the verification phase they could be facilitated in responding promptly to the robot attention request. Another possibility is that, in the RI condition, the robot is interacting in a more "active" way, because it asks questions about the objects. As observed by [14], this pro-active behavior regulates the interaction and provides a feedback signal to the human about the internal state of the robot. This behavior is also likely to induce in humans a social parenting effect: humans could have the impression that they are teaching the objects properties to a curious child. Conversely, in the HI case the robot acts "passively": it asks the human to provide the attention stimulus. So not only the learning process is led by the human, but the human could also be more hesitating in front of such request. The active/passive attitude could be responsible for making the robot more transparent to the human, in a way that the human would have or not a clear intuition about the robot's internal state. This claim is partially supported by some negative evaluations provided by the participants in the post-experiment questionnaire (see [9]). To summarize, the prior experience of an "active" robot leading the learning task makes the human react faster to the robot's attention utterances. Among the possible reasons, the robot active attitude improves its readability and the intuition of the human about the robot's state, hence the human reacts faster when he is interrogated by the robot. Our observations can be put in relation with the ones of [15], where they showed that "a robot responding to joint attention is more transparent, such that interactive task performance is

[1] In the experiment, the robot was programmed in a way that it was waiting for the human to match its referential focus. So an erroneous situation -i.e., the human looking at a different object than the one chosen by the robot- would have been caught. However, during the experiments this situation never occurred: the human always looked at the correct object pointed by the robot.

faster and more efficient". It is clear that the variable proactive element in the two conditions was decisive for setting the rhythm of the interaction in the two phases, especially in the verification phase. Though preliminary, our study suggests that the leading role of the robot can influence the pace of the interaction in a social task. Further, in this work we investigated also how social interaction can be made not only effective for the accomplishment of the task (i.e., teaching something to the robot) but also natural as if the human was teaching to another human. A remarkable result that suggest a natural interaction is the average amount of time that the participants in our study spent looking at the robot. According to [16], during human-human interaction people look at each other about 60% of the time, and look more while listening than while talking (during which they give frequent short glances). Interestingly the participants in our study looked at the robot about $65-67\%$ of the time, which is comparable with the human-human case. This could be a positive indicator of natural interaction[2]. Though preliminary, our results provide insights for improving the engagement system of the robot and make interaction with the human more natural and effective.

5. REFERENCES

[1] Wen Dong, B. Lepri, F. Pianesi, and A. Pentland. Modeling functional roles dynamics in small group interactions. *IEEE Transactions on Multimedia*, 15(1):83–95, 2013.

[2] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 383–388, 2005.

[3] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen. Interpersonal synchrony : A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, 3(3):349– 365, 2012.

[4] Charles Rich, Brett Ponsler, Aaron Holroyd, and Candace L Sidner. Recognizing engagement in human-robot interaction. In *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction (HRI)*, pages 375–382, 2010.

[5] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. Common metrics for human-robot interaction. In *ACM SIGCHI/SIGART Conference on Human-robot Interaction*, pages 33–40, 2006.

[6] A.L. Thomaz and C. Chao. Turn-taking based on information flow for fluent human-robot interaction. *AI Magazine: Special Issue on Dialog with Robots*, 32(4):53–63, 2011.

[7] Mark ter Maat, Khiet P. Truong, and Dirk Heylen. How turn-taking strategies influence users' impressions of an agent. In Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 441–453. Springer Berlin Heidelberg, 2010.

[8] MArk ter Maat, K. P. Truong, and D. Heylen. How agents' turn-taking strategies influence impressions and response behaviors. *Presence*, 20(5):412–430, 2011.

[9] S. Ivaldi, S.M. Anzalone, W. Rousseau, O. Sigaud, and M. Chetouani. Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. *Frontiers in Neurorobotics*, 8(5):DOI 10.3389/fnbot.2014.00005, 2014.

[10] Chen Yu, M. Scheutz, and P. Schermerhorn. Investigating multimodal real-time patterns of joint attention in an hri word learning task. In *ACM/IEEE Int. Conf. Human-Robot Interaction (HRI)*, pages 309–316, 2010.

[11] S. Ivaldi, S.M. Nguyen, N. Lyubova, A. Droniou, V. Padois, D. Filliat, P.-Y. Oudeyer, and O. Sigaud. Object learning through active exploration. *IEEE Transactions on Autonomous Mental Development*, page 10.1109/TAMD.2013.2280614, 2013.

[12] W. Rousseau, S. M. Anzalone, M. Chetouani, O. Sigaud, and S. Ivaldi. Learning object names through shared attention. In *Int. Conf. on Intelligents Robots and Systems - Workshop on Developmental Social Robotics*, 2013.

[13] 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 Neurorobotics*, 6:1–11, 2012.

[14] C. Breazeal. Toward social robots. *Robotics and Autonomous Systems*, 42:167–175, 2003.

[15] Chien-Ming Huang and Andrea Lockerd Thomaz. Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In *Proc. IEEE RO-MAN*, pages 65–71, 2011.

[16] Michael Argyle and Mark Cook. *Gaze and mutual gaze.* Cambridge U Press, 1976.

[2]Whether this value is affected by the embodiment induced by the humanoid shape, it is hard to tell: further experiments with different robots are necessary to investigate this question.