I reach faster when I see you look: Gaze effects in human-human and human-robot face-to-face cooperation

Jean-David Boucher, Ugo Pattacini, Amélie Lelong, Gérard Bailly, Frédéric Elisei, Sascha Fagel, Peter Ford Dominey, Jocelyne Ventre-Dominey

frontiers in
NEUROROBOTICS

ORIGINAL RESEARCH ARTICLE
published: 03 May 2012
doi: 10.3389/fnbot.2012.00003

I reach faster when I see you look: gaze effects in human–human and human–robot face-to-face cooperation

***Jean-David Boucher*[1]*, Ugo Pattacini*[2]*, Amelie Lelong*[3]*, Gerrard Bailly*[3]*, Frederic Elisei*[3]*, Sascha Fagel*[3]*, Peter Ford Dominey*[1]* and Jocelyne Ventre-Dominey*[1]**

[1] *Robot Cognition Laboratory, SBRI INSERM U846, Université de Lyon, Lyon, France*
[2] *Cognitive Humanoid Laboratory, Robotics, Brain and Cognitive Sciences Department, Istituto Italiano di Tecnologia, Genova, Italy*
[3] *GIPSA-Lab, UMR 5216 CNRS/INPG/UJF/U. Stendhal, Grenoble, France*

Edited by:
Alois C. Knoll, Technische Universität München, Germany

Reviewed by:
Anna M. Borghi, University of Bologna, Italy
Giovanni Ottoboni, University of Plymouth, UK

Correspondence:
Peter Ford Dominey, Robot Cognition Laboratory, INSERM U846, Université Lyon 1, 18 Avenue Doyen Lepine, 69675 Bron Cedex, France.
e-mail: peter.dominey@inserm.fr

Human–human interaction in natural environments relies on a variety of perceptual cues. Humanoid robots are becoming increasingly refined in their sensorimotor capabilities, and thus should now be able to manipulate and exploit these social cues in cooperation with their human partners. Previous studies have demonstrated that people follow human and robot gaze, and that it can help them to cope with spatially ambiguous language. Our goal is to extend these findings into the domain of action, to determine how human and robot gaze can influence the speed and accuracy of human action. We report on results from a human–human cooperation experiment demonstrating that an agent's vision of her/his partner's gaze can significantly improve that agent's performance in a cooperative task. We then implement a heuristic capability to generate such gaze cues by a humanoid robot that engages in the same cooperative interaction. The subsequent human–robot experiments demonstrate that a human agent can indeed exploit the predictive gaze of their robot partner in a cooperative task. This allows us to render the humanoid robot more human-like in its ability to communicate with humans. The long term objectives of the work are thus to identify social cooperation cues, and to validate their pertinence through implementation in a cooperative robot. The current research provides the robot with the capability to produce appropriate speech and gaze cues in the context of human–robot cooperation tasks. Gaze is manipulated in three conditions: Full gaze (coordinated eye and head), eyes hidden with sunglasses, and head fixed. We demonstrate the pertinence of these cues in terms of statistical measures of action times for humans in the context of a cooperative task, as gaze significantly facilitates cooperation as measured by human response times.

Keywords: human–human interaction, human–robot interaction, gaze, cooperation

INTRODUCTION

One of the most central and important factors in the real-time control of cooperative human interaction is the use of gaze (i.e., the combined orienting movements of the eyes and head) to coordinate and ensure that one's interlocutor is present, paying attention, attending to the intended elements in the scene and checking back on the status of the situation (Kendon, 1967). In this context, gaze is highly communicative both in indicating one's own attentional focus and in following that of the interlocutor. The importance of gaze is revealed in the specialization of brain systems dedicated to these functions (Perrett et al., 1992; Puce et al., 1998; Langton and Bruce, 1999; Langton et al., 2000; Pourtois et al., 2004; Calder et al., 2007).

Today, humanoid robots are of sufficient mechatronic sophistication that human–robot cooperation has become physically possible (Lallee et al., 2010a,b). We have observed that one of the current roadblocks in the elaboration of smooth and naturalistic human–robot cooperation is the coordination of robot gaze with the ongoing interaction. The objective of the current research is to identify pertinent gaze cues in human–human cooperative interaction, and to then test the impact of these cues in human–robot cooperation.

In order to analyze the role of gaze on human physical cooperation, we developed an experimental paradigm in which two subjects interact in order to identify and manipulate objects in a shared space. The paradigm has been designed such that the experiments are considered naturalistic and ecological by naïve subjects, but are sufficiently constrained to allow a rigorous behavioral psychology methodology, and robot implementation. The data obtained in the human–human experiment enabled us to determine how gaze is used (where, what, and when the subjects watch) in this cooperation task.

Based on these human gaze strategies, we created a simple model of task-related gaze control. This model has been implemented to control the gaze of the iCub humanoid. The iCub is a 53 degree of freedom humanoid robot with the body size of a 3-to 4-year-old child. It was developed in the context of a European project (FP6 RobotCub) as a European platform for the study of cognitive development. See http://icub.org/. The iCub was specifically developed to study embodied cognition, and thus has a highly articulated body, including a binocular oculomotor system that is physically capable of producing human-like oculomotor behavior.

We thus developed a new experiment in which one of the subjects was replaced by the robot. The setup and the instructions were based on those from the human–human interaction. As in human interactions, gaze proves to be a major clue in the interactions involving an embodied conversational humanoid.

THE IMPORTANCE OF GAZE

In face-to-face human communication, important non-verbal cues may accompany speech, appear simultaneously with speech, or even appear without the presence of speech at all. The most obvious non-verbal cues during speech communication originate from movements of the body (Bull and Connelly, 1985), the face (Collier, 1985), the eyes (Argyle and Cook, 1976), the hands and arms (Kendon, 1983), and the head (Hadar et al., 1983). More recent reviews can be found in (Pelachaud et al., 1996; McClave, 2000; Maricchiolo et al., 2005; Heylen, 2006). Gaze is linked to the cognitive and emotional state of a person and to the environment, hence, approaches to gaze modeling must address high level information about the communication process (Pelachaud et al., 1996; Lee et al., 2007; Bailly et al., 2010).

One of the central roles of such communicative processes is to allow the negotiation and on-line monitoring of cooperative behavior between individuals (Tomasello, 2008). A principal feature of cooperation is the creation and manipulation of a shared plan between the two cooperating agents. The shared plan is considered to provide a "bird's eye view" such that it includes the overall shared goal, and the breakdown in terms of "who does what, when" for both agents (Tomasello et al., 2005; Tomasello, 2008). Warneken et al. (2006) demonstrated the crucial role of gaze and other communicative acts in coordinating cooperation in the early stage of the development.

Recent research has made progress in the development of a shared plan capability in the context of human–robot cooperation (Lallee et al., 2010a,b; Dominey and Warneken, 2011). Dominey and Warneken demonstrated the ability of a robotic system to learn and exploit shared plans with a human partner. Lallee et al. (2010a,b) extended this work, providing the iCub robot the capability to watch two humans perform a cooperative task and to use vision to detect actions, agents, and goals, in order to create and use a shared plan describing the cooperative action. That shared plan can then be used by a humanoid to participate in achieving the shared goal, taking on either of the two possible roles (Lallee et al., 2010a,b). There is an important limitation in this latter work however, which is related to the lack of ongoing control of the cooperation by using and monitoring of gaze.

When humans interact in a shared environment, gaze can also relate to objects or locations in that environment that make up the object of joint attention, the third component of the "triadic" relation (Tomasello et al., 2005), and hence deliver information about the person's relation to that environment. While hand movements are often explicitly used to point to objects, head motion, and gaze yield implicit cues to objects (multimodal deixis, see Bailly et al., 2010) that are in a person's focus of attention as the person turns and looks toward the object of interest. Shared gaze is a natural and efficient method of communication that can significantly improve cooperative performance (Neider et al., 2010). Indeed, in tasks requiring rapid communication of spatial information, gaze may be more efficient than speech.

Gaze unveils the course of our cognitive activities, notably speech planning and spoken language comprehension. During interactive conversation, eye-reading is unconsciously used by interlocutors and boosts performance of the interaction. Chambers et al. (2002) have shown that linguistic and nonlinguistic information sources are combined to constrain referential interpretation and limit attention to relevant objects of the environment (see also Ito and Speer, 2006 for the integration of gaze and prosody).

In this context Hanna and Brennan (2007) examined gaze in a face-to-face cooperation task where one partner (the director) communicated the label of an object to be manipulated by the other partner (the matcher). In critical conditions the label for the target object was ambiguous, and had to be completed before the matcher could resolve the ambiguity via language. In these cases, the eye gaze produced by the director indicated the correct target before the verbal disambiguation, and crucially, this gaze was used by the matcher to resolve the temporary ambiguity. That is, before the verbal disambiguation was completed, the matcher had already directed his gaze to the correct target, guided by the gaze of the director.

Such results have now been extended into the domain of human–robot interaction. Staudte and Crocker (2009) exposed subjects to videos of a robot gazing at different objects in a linear array tangent to the line of sight, and sentences referring to these object, that were either congruent or incongruent with the videos. Subjects were to respond whether the sentence accurately described the scene. The principal findings of the study is the effect of robot gaze on human performance, with most rapid performance for congruent gaze, poorest performance for incongruent gaze, and intermediate performance when the robot made no gaze.

These studies indicate that humans pay attention to gaze (human and robot) and tend to follow it and use it under different task conditions. Hanna and Brennan (2007) demonstrate that humans' gaze is guided to the correct target by the gaze of their partner in a cooperative task. Staudte and Crocker (2009) demonstrate that such gaze effects can be produced by the gaze of a dual eyed robot. An open question that remains, is whether these effects of robot gaze can be generalized to performance improvements in physical interaction.

The present paper describes an experimental scenario of a face-to-face physical interaction in a shared environment. The accessibility of gaze information during the referencing of an object in the environment is manipulated and it is hypothesized that the task completion time increases when gaze cues are prevented or not perceived.

INTERACTION PARADIGM

The paradigm involves a cooperative task with two agents. The general idea is that the two agents each have partial knowledge, and they must cooperate and share knowledge in order to complete the task. One subject is called the informer, and the other the manipulator. The informer can see the target block and tell the manipulator where it is, but only the manipulator can move the

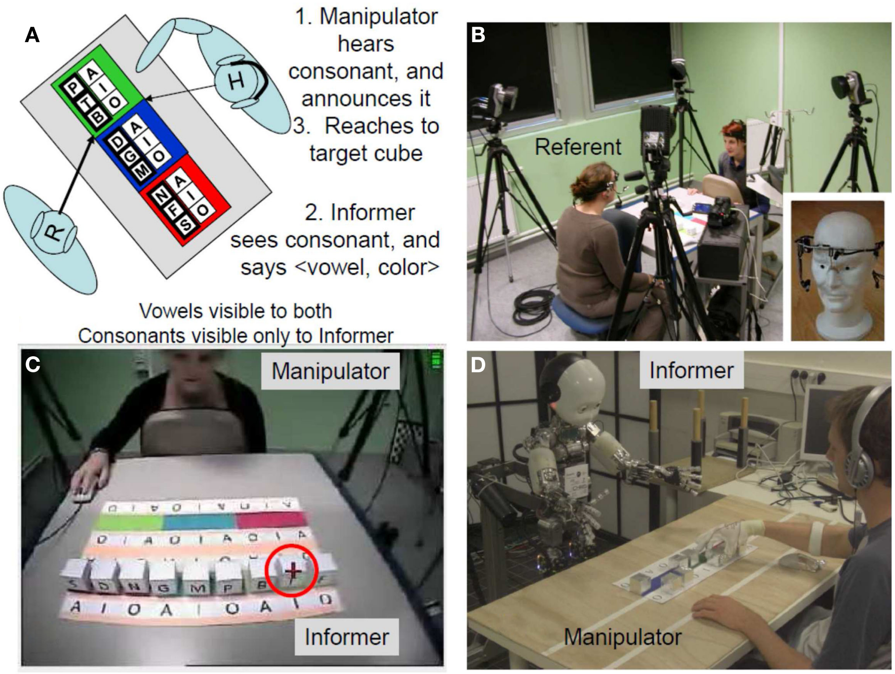
A
1. Manipulator hears consonant, and announces it
3. Reaches to target cube
2. Informer sees consonant, and says <vowel, color>
H
R
P A T I B O
A D I G O M
A N I F O S
Vowels visible to both
Consonants visible only to Informer
B
Referent
C
Manipulator
Informer
D
Informer
Manipulator

FIGURE 1 | Cooperation paradigm for human–human and human–robot interaction. (A) Schematic representation. Cubes labeled with consonants are on the playing surface. The consonants are visible only to the informer. The manipulator hears a consonant over headphones, and announces this to the informer. The informer performs a gaze search for the consonant cube, fixates it, and announces the vowel-color location to the manipulator who grasps the cube and puts in front of himself. **(B)** Human–human setup. Eye movements of the referent subject are recorded by eye tracker (see inset). **(C)** View from the informer, taken from the eye tracker. Circled red cross indicates current eye position. **(D)** Human–robot setup. The iCub plays the role of the informer.

block to achieve the shared goal. The informer and manipulator sit at opposite sides of a table as illustrated in **Figure 1**. On the table is a game board where the shared task is executed. The board is made up of three colored panels (red, blue, green from left to right), and on each of these colored panels are three slots labeled with vowels (A, I, O) as indicated in **Figure 1A**. Nine cubes are placed randomly in the slots of one of the two areas. Each cube shows a label from the set of consonants (P, T, B, D, G, M, N, F, S), which is only visible to the informer. The experiment can be described as follows:

(1) Via headphone, the computer confidentially tells the manipulator the consonant label of one cube to be moved. (The manipulator cannot see the consonant labels that he/she hears).
(2) The manipulator says the label to the informer in order to request the position of that cube [**Figure 2** (a)].
(3) The informer hears the consonant label [**Figure 2** (b)], and then searches among the cubes to find the one with that label [**Figure 2** (c)]. Note that after each six-move round, the cubes are pseudo-randomly reorganized, so the informer cannot memorize their locations.
(4) The manipulator waits for the informer's instruction [**Figure 2** (d)].
(5) The informer tells the manipulator the position of the requested cube by saying the vowel (A, O, or I) and color (red, green, or blue) of the slot where it is located [**Figure 2** (e)].
(6) The manipulator lifts her/his hand from the initial point [**Figure 2** (f)] and touches the cube.
(7) Once the cube is touched, the manipulator moves it into the corresponding slot (placed 10 cm in front) and puts back his hand to the initial position [**Figure 2** (g)].
(8) One round of the game consists of six such turns (i.e., cube identification and manipulation). After each round, the cubes are pseudo-randomly re-ordered on the playing surface.

We hypothesize that the manipulator can exploit the gaze of the informer in order to identify the target block more rapidly than if only the informer's spoken indication of the location was used. Thus in **Figure 2** (e) at the point where the informer begins to say the cube location, his gaze is already directed there, and can be exploited by the manipulator. Crucially, in this case, if the manipulator can indeed exploit this early gaze, then he may initiate the reaching movement before the end of the informer's sentence. In this case, the reaction time (RT; measured from the end of that sentence) can actually become negative. In order to experimentally control the visibility of informer's gaze by the manipulator, a condition is introduced in which the informer wears dark sun glasses, thus making vision of the eyes impossible for the manipulator.

HUMAN–HUMAN INTERACTION

HHI EXPERIMENTAL CONDITIONS

We monitored interactions between the referent subject and five naïve experimental subjects. The referent subject was equipped

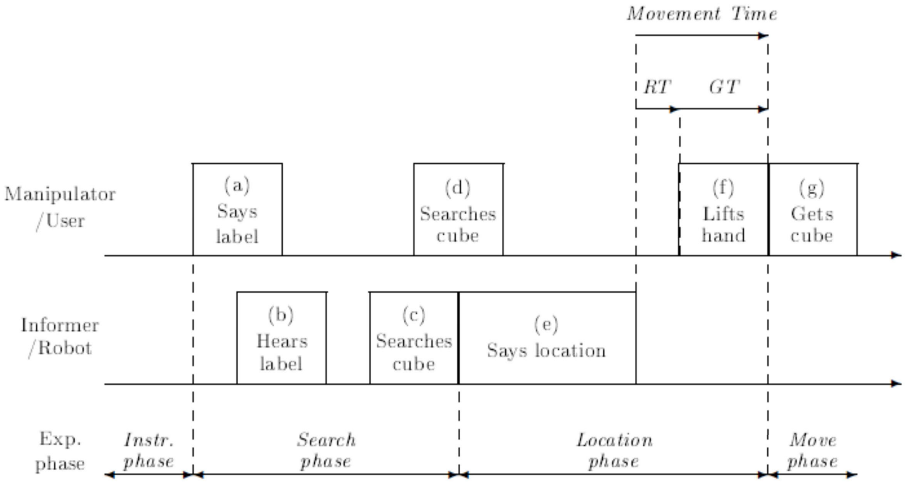

FIGURE 2 | Common timeline of the experiments. The task is composed of four phases. Instruction phase: from the confidential instruction provided by the program (via the headphone to the manipulator) to the beginning of the label verbalization (a). Search phase: from the beginning of the label verbalization by the manipulator (a), the informer hearing (b), and then searching for (c) the cube, to the beginning of the location verbalization by the informer (e). The location phase: from the label verbalization by the informer (e) to manipulator's hand contact on the cube (f). Move phase: the rest of the hand movement, from the cube contact to the initial position. RT is the reaction time, from the end of Says location (e) to the beginning of movement onset (f). Grasp time (GT) is from movement onset, to contact with the cube. Movement time is the sum RT and GT. Note that if the manipulator lifts the hand (f) before the end of Says location (e), then the RT will be negative, i.e., anticipatory.

Table 1 | Experimental conditions.

Condition	Referent subject (with eye tracker)	Experimental subject	Note
1. MI	Manipulator	Informer	
2. IM	Informer	Manipulator	
3. MIg	Manipulator	Informer with glasses	Predict effects on manipulator From informer wearing glasses
4. IMg	Informer with glasses	Manipulator	Predict effects on manipulator From informer wearing glasses

Experimental subject is informer in 1 and 3, and manipulator in 2 and 4. To test the effect of hidden informer eye position, the informer wears glasses in conditions 3 (experimental subject) and 4 (reference subject) thus preventing manipulator from seeing informer gaze.

with a head mounted eye tracker in all conditions (see technical description below, and **Figure 1B**). During each interaction the referent subject acted as manipulator in six rounds and as informer in another six rounds. She (the referent subject) wore sun glasses in half of the rounds and did not wear sun glasses in the other half. All rounds are grouped in blocks with the same role assignment and condition (sun glasses or not). The order of these blocks was counterbalanced across the four conditions defined in **Table 1**.

Two training rounds of three turns were played before the recording (one for each role assignment) and the subject was instructed to play fast but accurately. We randomized the cubes positions at the beginning of each round. Thus, in conditions 1 and 2 the manipulator can see the informer's eyes, and in conditions 3 and 4 the informer's eyes are hidden from the manipulator by the glasses.

HHI TECHNICAL SETUP

During the interaction we recorded: (i) both subjects' head motions with an HD video camera (both subjects simultaneously by using a mirror), and by a motion capture system, (ii) the subjects' speech, by head mounted microphones, and (iii) the gaze of the referent subject and a video of what she sees by a head mounted eye tracker (Pertech eye tracker; see http://www.pertech.fr; **Figure 1B** inset). Note that the eye tracker functions with infrared light: it was not affected by the sun glasses. We also monitored the timing of the moves by the log of the script that controls the experiment. The different data streams were post-synchronized by recording the sync signal of the motion capture cameras as an audio track along with the microphone signals as well as the audio track of the HD video camera, and by a clapper board that is recorded by the microphones, the scene camera of the eye tracker and the motion capture system simultaneously. **Figure 2** shows an overview of the technical setup.

MEASUREMENTS AND LABELING

The timing of the computer generated events and subjects' responses are imported from the log of the control script. Additional cues characterizing subjects' behaviors are collected semi-automatically in the following way. Speech instructions uttered by each member of the dyad are annotated with Praat (Boersma and Weenink, 2007). Gaze fixations were identified in the raw gaze data (from the eye tracker worn by the referent subject) using a dispersion-based algorithm (Salvucci and Goldberg, 2000). These fixations are further attributed by hand in ELAN (Berez, 2007; Hellwig et al., 2010) to one of 24 different regions of interest (ROI): One of the nine cubes, one of the nine target areas, the mouth, left or right eye of the subject, the mouse, the hand of the manipulator, and elsewhere.

The timing of the four different phases of the interaction (**Figure 2**) are derived from these data as follows:

(1) Instruction phase: From onset of the confidential instruction provided by the computer to the manipulator, until the start of the verbalization of the cube label (consonant) by the manipulator.
(2) Search phase: From consonant cue label verbalization onset until onset of the verbalization of the cube location (vowel, color) by the informer. This is the time needed by the informer to search for the cube corresponding to the randomly chosen label and to get the corresponding location.
(3) Location phase: From onset of target location verbalization until the end of the manipulator's lifting. This is the time needed by the manipulator to locate the cube and to begin the grasping movement. The phase is composed by two periods; the location speech signal and the movement time (MT), and terminates when the manipulator touches the cube. This is the critical period where manipulation of informer gaze should affect the manipulator's RT.
(4) Move phase: From the first hand-cube contact to the hand return to the initial position.

HHI RESULTS

We analyzed the impact of gaze on the durations of these four phases of the interaction (**Figure 3**). We were particularly interested in determining which components of the task are impaired when the manipulator's access to informer's gaze is obstructed by the informer wearing glasses.

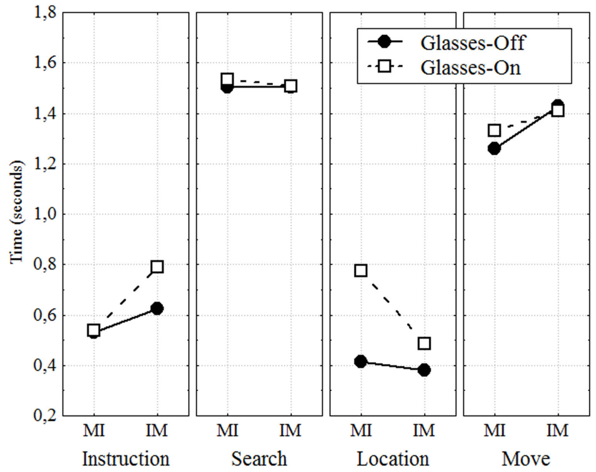

FIGURE 3 | Gaze Effects. Distributions of durations of the four different phases of the interaction with gaze blocked by Glasses (glasses-on), or not (glasses-off). During the Location phase when the Informer is wearing glasses, the manipulator's response time is significantly impaired.

Gaze effects

Statistical analysis was realized by using repeated measures ANOVA with the software package Statistica. The dependent variable is completion time for the different phases of the task. The independent variables are Phase, Glasses (with and without), and Role (referent as manipulator and subject as manipulator) to focus on the effect of hiding the informer's gaze from the manipulator. *Post hoc* analysis was provided by planned comparisons. The significance level was established at a 95% confidence interval.

Figure 3 illustrates the effects of sun glasses on the durations of the four experimental phases. Examining the four phases, comparing the glasses-on to glasses-off, it appears that the glasses effect is strongest in the Location phase. The differential effect of the glasses-on duration performance in each phases was 86 ms (Instruction), 19 ms (Search), 233 ms (Location), and 26 ms (Move). Thus, the 233 ms effect in location phase was by far the largest. This observation was confirmed by the repeated measures ANOVA. Overall there was a small and non-significant main effect of glasses [$F(1, 4) = 5.5$, $p = 0.08$] with durations with glasses at 1.04 s, and without at 0.95 s. The Role effect was highly non-significant [$F(1, 4) = 0.6$, $p = 0.5$] indicating that the mapping of referent and subject to informer and manipulator had no influence on performance. The Phase effect was highly significant [$F(3, 12) = 42.1$, $p < 0.0001$] simply reflecting the fact that the different phase have different durations. Most important, the only interaction that approached significance was the Glasses × Phase [$F(3, 12) = 3.4$, $p = 0.05$]. Planned comparisons revealed that the informer wearing glasses had a significant effect only in the Location phase with $p < 0.001$. In the other three phases the glasses had no effect ($p > 0.1$). This confirms that the manipulator relies on informer gaze in order to locate and reach to the target cube. When informer's gaze is blocked, the manipulator's location time is significantly increased.

Gaze patterns by phase

Thus, visibility of the informer's gaze influences the manipulator. To better understand this effect, we examined where the manipulator and informer are looking during the interaction. **Figure 4** illustrates the gaze behavior of the referent subject in the four conditions (columns), during the four task phases (rows). Recall that in different blocks, the referent subject plays the role both of the informer and manipulator, respectively. The consistent behavior of gaze during the instruction and location phases (rows 2 and 4 from the top of **Figure 4**) for both roles illustrates their mutual attention for common ground, i.e., for the face during instruction and target regions during location. The distribution of gaze during the crucial Search phase (row 3) depends strongly on role: while the informer searches for the correct target, the manipulator follows the informer's gaze, looking at his eyes, and the game board locations (source and target).

HHI DISCUSSION

The results of this experiment allow us to make two significant conclusions. The informer indeed signals the location of the target cube by gaze, and the manipulator exploits this gaze in making the reaching movement toward the target block more rapidly. This confirms and extends the work of Hanna and Brennan (2007).

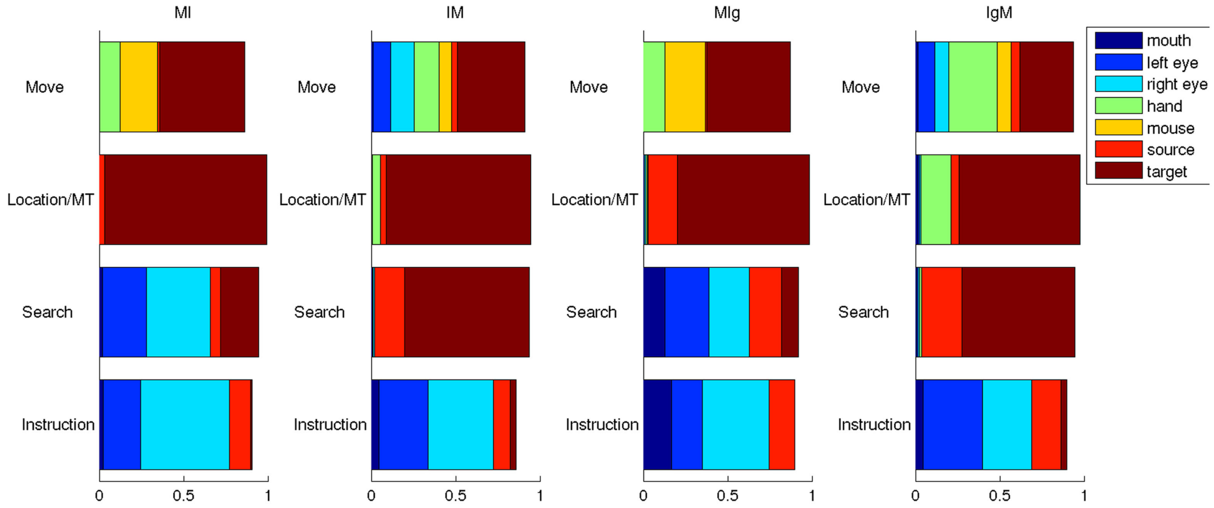

FIGURE 4 | Distributions of referent subject's gaze toward ROIs according to phase (rows) and condition (columns). During search, Manipulator looks at Informer's eyes, while Informer looks at target blocks to identify the named cube.

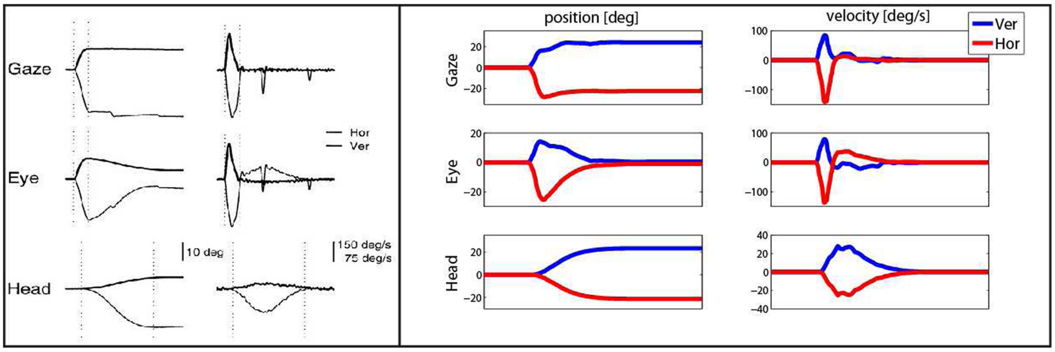

FIGURE 5 | Qualitative comparison between human (left) and iCub (right) saccadic responses: position (first column) and velocities (second column) of the head and the eye movements (both horizontal and vertical) are given when the human and the robot are asked to look at the target. In both cases, the first row represents the gaze response as result of the coordinated rotation of the head and the eye, which are illustrated in the second and third rows, respectively.

They demonstrated that the manipulator's gaze follows that of the informer, and we demonstrated that this allows for more efficient behavior in the cooperative task. As illustrated in **Figure 4**, during the search phase, while the informer is looking for the target block, her gaze is principally directed toward the target blocks. During the same period, the manipulator scrutinizes the informer's eyes, apparently trying to "read" the gaze to identify the final target block upon which the informer's gaze will fall. Finally, the results with the sun glasses bear this out. In the presence of sun glasses, the duration of the location phase is significantly increased, as illustrated in **Figure 3**.

HUMAN–ROBOT INTERACTION EXPERIMENT

The results of the human–human experiment allows us to hypothesize that when the informer's gaze can be seen by the manipulator, the manipulator can use this gaze information to identify the target cube and start the successful reach to that cube significantly faster than when gaze is blocked. In order to test this hypothesis, we should create a situation in which the robot acts as informer, in conditions in which its gaze is vs. is not visible to the human subject playing the role of manipulator (Boucher et al., 2010).

HRI TECHNICAL SETUP

The human–robot interaction takes place with the iCub. In the interaction, the iCub is always the informer. The robot is seated 50 cm from the aligned cubes on the playing surface (see **Figure 5**), with the manipulator (naïve human subject) sitting across the table. The manipulator confidentially hears the specification of one of the consonant cube labels and then repeats it to the robot. While the manipulator announces the label, the robot is looking at the manipulator. One second after the end of the speech detection (of the manipulator repeating the label), the robot looks away from the manipulator's face, and begins to search for the specified cube. We program the gaze to the target cube to pseudo-randomly take between 1 and 3 saccades, alternating between the left and right side, to the remaining cubes in the display, approximating the ocular search behavior observed in the human subjects. The robot's eye movement and head movement completion times are respectively 100 and 600 ms. The eye thus attains the target first, with the head continuing to move to the target, and the eyes compensating for this continued head movement in order to stay fixated on the target.

This way the robot is able to realistically mimic the well known behavior human subjects show when asked to perform rapid eye saccades. **Figure 6**, indeed, points out the high level of agreement between data recorded from experiments on human (Goossens and Van Opstal, 1997) and the traces of head and eyes displacements obtained from iCub during gazing tasks: the eyes quickly rotate toward the target (saccades) and then start counter-rotating to compensate for the slower neck movement during the vestibular driven phase in order to maintain the gaze stable fixation until the head comes to rest.

The generation of these human-like movements studied in human gaze is achieved in the robot with the iCub gaze controller. As described in (Pattacini, 2010), the controller employed to coordinate the iCub gaze acts intrinsically in the Cartesian space, taking as input the spatial location of the object of interest where to direct the robot attention, and then generates proper velocity commands simultaneously to the neck and the eyes. Therefore the task boils down to finding a suitable configuration of the head and eyes joint angles that allows the 3D robot fixation point to coincide with the desired input location. The inverse problem is tackled with a pair of solver-controller components, where the solver is responsible to identify the final configuration with the resort to a sophisticated – yet fast – non-linear constrained optimization (Wätcher and Biegler, 2006), whereas the controller is in charge of generating velocities with minimum-jerk profiles that steer the parts toward the solved state. Minimum-jerk velocity profiles have the interesting property of being bell-shaped, entailing a well defined onset and completion of the trajectory; it has been discovered that this property is widely present in human movement.

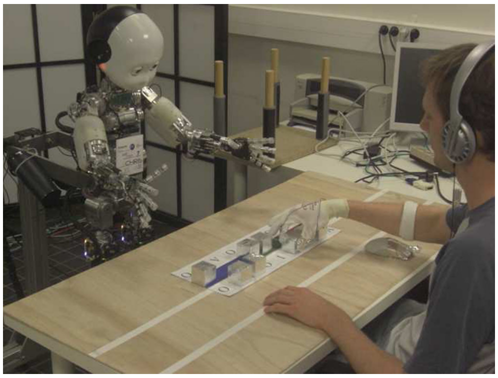

FIGURE 6 | The HRI setup. On the left side, the robot plays the role of the informer whereas the human is the manipulator. According to the conditions, the robot changes its behavior (see **Figures 7A,C**). In order to observe a potential anticipation due to the eye and/or the head movement we measure the reaction time (RT) and the movement time (MT). The RT is the duration between the end of the speech location signal (**Figure 2** part (e)) and when the human lifts her/his hand from the initial position (on the picture under the elbow). The MT is the duration between the end of the speech location signal and the first manipulator cube contact (details in **Figure 7D**). See **Figure 1** for timeline.

The design envisages two independent blocks, each composed of solver-controller units: one devoted to the neck control and one for handling the eyes. Finally, the gyroscope mounted within the iCub head allows collecting the vestibular data that are required to compensate the head rotation at the eyes level. Once the final eyes and head position are sent to the robot, it waits for a random delay ranges from 500 to 1000 ms, and announces the location of the cube.

In order to accurately measure the manipulator's manual performance, we developed a novel behavior recording device. A USB mouse was disassembled, and the left and right click buttons were replaced by mechanical/electrical contacts (see **Figures 6** and **7D**). One set of contacts is fixed on the index finger and thumb of a glove that the manipulator wears on her/his right hand. When these contacts come into physical contact with a metallic cube, the event is registered on the USB device. Similarly, the two contacts from the left mouse button are connected to metal contacts on the palm of the glove, and resting position marker on the playing surface. This allows precise detection of the onset of the movement to reach for the cube (RT), when the palm leaves the start position (see **Figure 2** for timeline).

HRI EXPERIMENTAL CONDITIONS

We studied the effects of robot (informer) communicative behavior on the naïve human subjects in three conditions. In all cases, the informer (the iCub) provides verbal information specifying

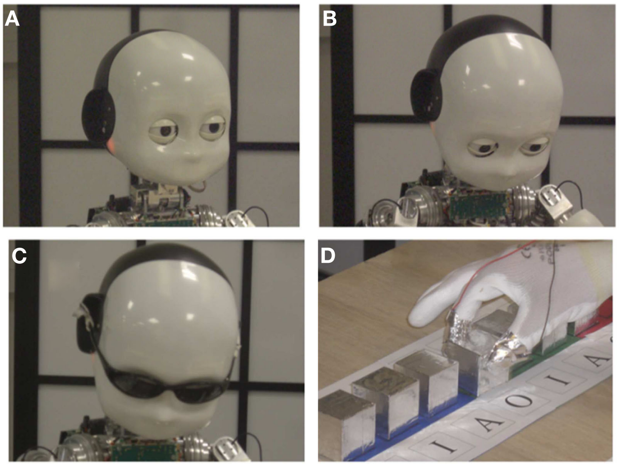

FIGURE 7 | Human–Robot interaction conditions (A–C) and the contact detection glove (D). In the head fixed condition **(A)**, the robot stays in a neutral position, moving neither the eyes nor head. In the full gaze condition **(B)**, the robot searches for the correct cube with coordinated gaze (eye and head movement). The sun glasses condition **(C)**, is the same as **(B)** except the robot wears sun glasses. In this condition, the human cannot see the eyes. **(D)** Illustrates the contact detection glove. Contact is detected when the electrical circuits are closed. There two kinds of contact detections: (i) when the glove touches/releases a cube and (ii) when the glove palm contacts/moves from the initial position.

the target color and vowel to the naïve manipulator. What is varied systematically across conditions is the amount of information provided by gaze, in three conditions: (1) Head fixed condition. This is only in the human–robot experiments. There is no head nor eye movement generated by the iCub informer (**Figure 7A**). Target cube position information is provided solely by speech. (2) Full gaze condition. The robot indicates the location of the target cube by a coordinated eye and head movement (**Figure 7B**), and by speech. (3) Sun glasses condition. The human manipulator cannot see the iCub's eyes as the iCub is wearing sun glasses (**Figure 7C**). Thus the initial positioning information provided by the eyes is not available. The manipulator must wait 600 ms for the completion of the head movement.

Note that an anticipatory motion of the hand to the initial position and toward the target can be predicted in the full gaze and sun glasses conditions, as the deictic gaze motion precedes the speech command. Critically, the gaze information becomes available before the end of the speech signal.

HRI PERFORMANCE MEASUREMENTS

To initiate a trial, the subject puts her/his palm on the resting position marker and this is recorded. As the trial proceeds, when the subject lifts her/his hand from the initial position this is recorded as the movement and RT onsets (see RT and MT on **Figure 2**). Contact with the cube is recorded to generate the MT. Finally grasping time (GT) is the duration from the RT offset to MT offset. In other words, the GT corresponds to the period in which the subject lifts her/his hand from the initial position to the first cube contact (**Figure 2F**). Five naïve subjects were exposed to 10 games or rounds in each of three conditions for a total of 30 rounds per subject, with the total duration of approximately 50 min, divided into two 25 min sessions. The three conditions were full gaze, sun glasses, and head fixed. The sun glasses and full gaze conditions are similar to the HHI conditions with or without sun glasses. The third condition (head fixed) is difficult or impossible to be performed by human subjects, who cannot inhibit their natural gaze toward the target cubes. The use of the iCub allows us to explore this condition.

HRI RESULTS

The results are presented in the context of the three experimental conditions, full gaze, sun glasses, and head fixed. In **Figure 8** we illustrate the two key performance measures: the RT, the MT.

Recall that the RT is characterized as the numerical difference in seconds between the end of the speech signal of the location (vowel, color) from the robot informer (**Figure 2E**), and the lifting of the human manipulator's hand from the initial position (**Figure 2F**).

It should be made clear that the robot's gaze can aid the manipulator to identify the target location *before* the target has been specified by speech. Thus, the human manipulator can potentially anticipate the speech location signal, especially in the full gaze and sun glasses conditions.

Again, the MT is characterized as the delay between the offset of the informer (robot) specification of the target (vowel, color) location, and the human subject's first touch of the corresponding cube. MT, RT, and GT are simply related by the expression: MT = RT + GT.

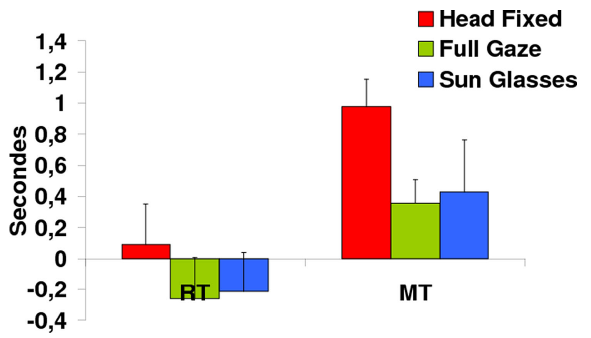

FIGURE 8 | Reaction time (RT) and movement (MT). RT is positive for Head Fixed, and negative (anticipatory) for Full Gaze and Sun Glasses, indicating that manipulator can read the iCub gaze. This anticipation is reflected in the movement time, as MT = RT + GT.

Statistical analysis was realized by using repeated measures ANOVA with the software package Statistica. The dependent variables were the RT, the MT, and the GT. The within subjects factors were the conditions (head fixed, full gaze, sun glasses), and rounds (from 1 to 10). *Post hoc* comparison was provided by Scheffe test analysis. The significance level was established at a 95% confidence interval.

As illustrated in **Figure 8** the three conditions (head fixed, sun glasses, and full gaze) had visible influence on both the RT [main condition effect: $F(2, 8) = 23$, $p < 0.001$] and MT [main condition effect: $F(2, 8) = 41$, $p < 0.001$]. The mean RT and MT were significantly different for head fixed vs. sun glasses and full gaze at $p < 0.01$ and $p < 0.01$ respectively (Sheffe *post hoc*).

In contrast, GT was independent of the conditions [main condition effect: $F(2, 8) = 1$, $p = 0.39$] which indicates that the experimental conditions influence the movement onset (RT) but not the movement trajectory (i.e., the delay between lifting the hand from the start position and touching the target block). This result confirms our hypothesis that the informer's eye and/or head movements can be used by the manipulator to anticipate the specification of the target cube location.

Figure 9 displays RTs over the 10 successive rounds. Interestingly, during the first two rounds, there is a clearly visible performance gradient with the head fixed condition yielding positive RTs, the sun glasses condition yielding marginally anticipatory (negative) RTs, and finally the full gaze condition yielding even greater anticipatory RTs. In subsequent blocks, the RTs for the sun glasses condition approach those for the full gaze condition.

When analyzing the mean RT over the rounds, we observed an effect of rounds (learning) mainly in the full gaze and sun glasses conditions: as illustrated in **Figure 9**, the learning effect was due to a decrease of mean RT over the rounds (Condition × round) interactions: [$F(1, 45) = 1.8$, $p = 0.02$]. The adaptation of the manipulator to the sun glasses condition is revealed as we observe a significant difference in the mean RT of full gaze and sun glasses conditions only in the two first turns ($p < 0.01$). This demonstrates that the effects of sun glasses is attenuated over the course of the 10 rounds.

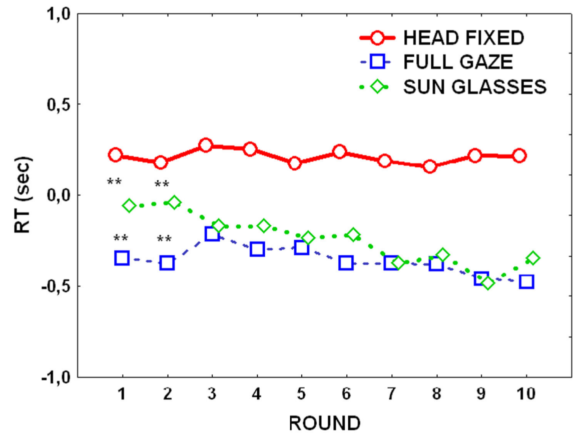

FIGURE 9 | Reaction time (RT) over the rounds. In initial rounds Head Fixed > Sun Glasses > Full Gaze. Later, this becomes Head Fixed > Sun Glasses = Full Gaze, indicating that subjects learn to "read" the gaze from head position of the robot.

HRI DISCUSSION

In this experiment, the robot's speech specification of the target cube is slightly preceded by its gaze motion to that target. We could thus predict that subjects will attend to this information and begin to move their hand toward the target before completion of the speech location signal, exploiting the information of the gaze signal. This corresponds to the negative (anticipatory) values for the RT in **Figure 8**, for the conditions in which gaze was present (with or without shielding of the eyes by the sun glasses). This advantage for the full gaze and sun glasses conditions was likewise transferred to the MT. The essentially sensorimotor aspects of the grasping motion itself were not influenced by our experimental manipulation of head and/or eye motion (the grasping time GT is independent of the conditions). This indicates that, once a subject lifted their hand from the initial position, the time to complete the movement was not affected by the type of the conditions.

A novel finding is that the sun glasses produce a significant impairment in the first two blocks, which is then attenuated. We note that the head position indeed provides the same information as the eye position, well before the end of the spoken target specification. It is thus likely that human subjects progressively learn to exploit this redundant information. Indeed, Hudson et al. (2009) demonstrate that eye and head movements are highly functionally correlated, and relations between them are used by humans in a predictive manner.

COMPARISON OF HUMAN–HUMAN AND HUMAN–ROBOT INTERACTION

As summarize in **Table 2**, we can compare the MT of the two experiments and also make the analogies between the IM and IgM conditions and the full gaze and sun glasses conditions. In the HHI (and in Fagel et al., 2010), we observe a significant sun glasses effect, i.e., when the manipulator could not see the informer's eyes, the duration of the Location phase was increased. In the HRI, we also observe a significant sun glasses effect, particularly in the first rounds of the interaction. Thus, the robot informer provides gaze information via eye and head position that can be exploited by the human manipulator. Interestingly, the sun glasses effect was subsequently diminished due to a potential implicit learning effect. Indeed, as illustrated in **Figure 9**, from the third round, RTs in the Glasses-On condition, approach those in the Full Gaze condition. It is likely that subjects implicitly learn to extract the target location from the head movements when the eyes are covered. Indeed, attentional orientation can be significantly decoded directly from gaze, and from head position without vision of the eyes (Ricciardelli et al., 2002, 2007, 2009; Ricciardelli and Driver, 2008). Further investing this with human–robot interaction is a topic for future research.

Table 2 | Comparison of contact detection systems, times measured, and conditions between the HHI and the HRI.

	Human–human interaction	**Human–robot interaction**
Detection of manipulator cube contact	Off line video analysis	Electromechanical contact
Movement time: from end of location speech to the hand-cub contact	MT	MT = RT + GT
Informer gaze visible	IM and MI conditions	Full gaze condition
Informer gaze hidden	IMg and MIg conditions	Sun glasses condition
Informer head fixed	None	Head fixed condition

GENERAL DISCUSSION

When do human subjects exploit spatially directed targets as if they were gaze cues? This question was addressed in the work of Jonides (1981). He conducted experiments in which two types of spatial cues were presented to subjects: (i) pictures of eyes, and (ii) arrows. He demonstrated that the eye pictures displayed in the center of a computer screen triggered reflexive shifts, whereas the arrows did not. More recently, Hanna and Brennan (2007) demonstrated that human subjects, the "matcher," and "director" exploit human gaze in disambiguating language. Before the verbal disambiguation was completed, the matcher had already directed his gaze to the correct target, guided by the gaze of the director. Staudte and Crocker (2009) have now extended such approaches into the domain of human–robot interaction. The principal findings of the study is the effect of robot gaze on human gaze performance, with most rapid gaze for congruent gaze, poorest performance for incongruent gaze, and intermediate performance when the robot made no gaze. An open question that remains, is whether these effects of robot gaze can be generalized to performance improvements in physical interaction.

The current research extends these results into the domain of physical interaction. In a human–human cooperative interaction, we observed human performance which led to the hypothesis that informer's gaze toward the target location could be used by the manipulator to produce faster RTs to the target of shared interest. We then tested the corresponding human–robot interaction where we tested this hypothesis. Our innovative results confirmed that indeed, naïve humans can reliably exploit robot gaze to allow them to perform in an anticipatory manner in a cooperative task. It should be noted that the gaze has coordinated eye and head movement whose dynamics are inspired by those of the human. Based on the results obtained in HHI and HRI, we extend the notion that the eye plays a significant role in human interactions and go a step further embodied experiments with interaction between naïve humans and a humanoid robot.

This research continues in the ongoing trajectory of studies on communicative human–robot interaction. Previous studies have performed detailed measurements of the effects of robot motion on human engagement (Sidner et al., 2005). We extend this in a complementary way to look at the effects of robot gaze on human performance in cooperative tasks.

Our results indicate that for robots with articulated eyes and head, such as the iCub, naïve human subjects are sensitive to gaze in the context of cooperative tasks. This is promising for the use of gaze in enabling affective human–robot interaction. In particular, part of the coordination of shared plans includes partners acknowledging that they have completed their turn. In the past we have done this explicitly, using language (Dominey and Warneken, 2011). The use of gaze as a communicative mechanism for coordinating cooperative shared plans could significantly increase the quality of these interactions.

The results obtained in our research clearly indicate that human subjects can effectively exploit the gaze cues of human and robot partners in a physical interaction task. Interestingly, neurophysiological data indicate that the primate brain has specific neural networks for recognition of gaze (Perrett et al., 1992; Puce et al., 1998; Langton and Bruce, 1999; Langton et al., 2000; Calder et al., 2007). These networks respond mainly to stimuli related to eye and head perception (and not to physical stimuli, e.g., arrows). It is now being established that, with the proper visual features and control dynamics, robotic, and avatar systems can be shown to recruit brain systems that are required in processing human social behavioral cues (Chaminade and Cheng, 2009).

Interestingly, more than attention can be communicated through perception of body language. de Gelder (2009) reviews distinct neurophysiological mechanisms for reading and interpreting emotional body language, suggesting that effective robot communication should include appropriate recruitment of the whole body. In addition to gaze, fully body posture can be used to communicate emotional and likely cognitive states (Meeren et al., 2005; de Gelder, 2006, 2009; van de Riet et al., 2009).

Future research will extend these results into the domain of human–robot interaction.

ACKNOWLEDGMENTS

This research has been supported by the French ANR Project AMORCES, and the FP7 project CHRIS and EFAA.

REFERENCES

Argyle, M., and Cook, M. (1976). *Gaze and Mutual Gaze*. Oxford: Cambridge University Press.

Bailly, G., Raidt, S., and Elisei, F. (2010). Gaze, conversational agents and face-to-face communication. *Speech Commun.* 52, 598–612.

Berez, A. L. (2007). Eudico linguistic annotator (Elan). *Lang. Document. Conserv.* 1, 283–289.

Boersma, P., and Weenink, D. (2007). *Praat: Doing Phonetics by Computer (Version 5.1.08) (Computer Program)*. Available at: http://www.praat.org/ [retrieved June 28, 2009].

Boucher, J.-D., Ventre-Dominey, J., Fagel, S., Bailly, G., and Dominey, P. F. (2010). "Facilitative effects of communicative gaze and speech in human-robot cooperation," in *ACM Workshop on Affective Interaction in Natural Environments (AFFINE)*, ed. G. Castellano (New York: ACM), 71–74.

Bull, P., and Connelly, G. (1985). Body movement and emphasis in speech. *J. Nonverbal Behav.* 9, 169–187.

Calder, A. J., Beaver, J. D., Winston, J. S., Dolan, R. J., Jenkins, R., Eger, E., and Henson, R. N. A. (2007). Separate coding of different gaze directions in the superior temporal sulcus and inferior parietal lobule. *Curr. Biol.* 17, 205.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., and Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *J. Mem. Lang.* 47, 30–49.

Chaminade, T., and Cheng, G. (2009). Social cognitive neuroscience and humanoid robotics. *J. Physiol. Paris* 103, 286–295.

Collier, G. (1985). *Emotional Expression*. Hillsdale: Lawrence Erlbaum.

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* 7, 242–249.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3475–3484.

Dominey, P. F., and Warneken, F. (2011). The basis of shared intentions in human and robot cognition. *New Ideas Psychol.* 29, 260–274.

Fagel, S., Bailly, G., Elisei, F., and Lelong, A. (2010). "On the importance of eye gaze in a face-to-face collaborative task," in *ACM Workshop on Affective Interaction in Natural Environments (AFFINE)* (Firenze), 81–85.

Goossens, H. H., and Van Opstal, A. J. (1997). Human eye–head coordination in two dimensions under different sensorimotor conditions. *Exp. Brain Res.* 114, 542–560.

Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1983). Kinematics of head movements accompanying speech during conversation. *Hum. Mov. Sci.* 2, 35–46.

Hanna, J., and Brennan, S. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *J. Mem. Lang.* 57, 596–615.

Hellwig, B., Uytvanck, D. V., and Hulsbosch, M. (2010). *ELAN – Linguistic Annotator. Manual, Version 3.9.0.* Nijmegen: Max Planck Institute for Psycholinguistics.

Heylen, D. (2006). Head gestures, gaze and the principles of conversational structure. *Int. J. HR* 3, 241–267.

Hudson, M., Liu, C. H., and Jellema, T. (2009) Anticipating intentional actions: the effect of eye gaze direction on the judgment of head rotation. *Cognition* 112, 423–434.

Ito, K., and Speer, S. R. (2006). "Immediate effects of intonational prominence in a visual search task," in *Proceedings of Speech Prosody*, Dresden, 261–264.

Jonides, J. (1981). "Voluntary versus automatic control over the mind's eye's movement," in *Attention and Performance IX*, Vol. 9, eds J. B. Long and A. D. Baddeley (Hillsdale: Erlbaum), 187–203.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol. (Amst.)* 26, 22–63.

Kendon, A. (1983). *Gesture and Speech: How they Interact*, Vol. 11. Beverly Hills: Sage Publications, 13–45.

Lallee, S., Lemaignan, S., Lenz, A., Melhuish, C., Natale, L., Skachek, S., Zant, T. V. D., Warneken, F., and Dominey, P. F. (2010a) "Towards a platform independent cooperative human-robot interaction system: I. perception," in *The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2010*, Taipei.

Lallee, S., Madden, C., Hoen, M., and Dominey, P. F. (2010b). Linking language with embodied and teleological representations of action for humanoid cognition. *Front. Neurorobot.* 4:8. doi:10.3389/fnbot.2010.00008

Langton, S. R., Watt, R. J., and Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends Cogn. Sci. (Regul. Ed.)* 4, 50–59.

Langton, S. R. H., and Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Vis. Cogn.* 6, 541–567.

Lee, J., Marsella, S., Traum, D., and Gratch, J. (2007). "The Rickel gaze model: a window on the mind of a virtual human," in *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA-2007). Lecture Notes in Computer Science*, Vol. 4722, eds C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis and D. Pelé, September 17–19, Paris, 296–303.

Maricchiolo, F., Bonaiuto, M., and Gnisci, A. (2005). "Hand gestures in speech: studies of their roles in social interaction," in *Proceedings of* *the Conference of the International Society for Gesture Studies*, Lyon.

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *J. Pragmat.* 32, 855–878.

Meeren, H. K., van Heijnsbergen, C. C., and de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16518–16523.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., and Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychon. Bull. Rev.* 17, 718–724.

Pattacini, U. (2010). *Modular Cartesian Controllers for Humanoid Robots: Design and Implementation on the iCub*. Ph.D. dissertation, RBCS, Istituto Italiano di Tecnologia, Genoa.

Pelachaud, C., Badler, N. I., and Steedman, M. (1996). Generating facial expressions for speech. *Cogn. Sci.* 20, 146.

Perrett, D., Hietanen, J., Oram, M., Benson, P., and Rolls, E. (1992). Organization and functions of cells responsive to faces in the temporal cortex (and discussion). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 23–30.

Pourtois, G., Sander, D., Andres, M., Grandjean, D., Reveret, L., Olivier, E., and Vuilleumier, P. (2004). Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals. *Eur. J. Neurosci.* 20, 3507–3515.

Puce, A., Allison, T., Bentin, S., Gore, J. C., and McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.* 18, 2188–2199.

Ricciardelli, P., Betta, E., Pruner, S., and Turatto, M. (2009). Is there a direct link between gaze perception and joint attention behaviours? Effects of gaze contrast polarity on oculomotor behaviour. *Exp. Brain Res.* 194, 347–357.

Ricciardelli, P., Bonfiglioli, C., Iani, C., Rubichi, S., and Nicoletti, R. (2007). Spatial coding and central patterns: is there something special about the eyes? *Can. J. Exp. Psychol.* 61, 79–90.

Ricciardelli, P., Bricolo, E., Aglioti, S. M., and Chelazzi, L. (2002). My eyes want to look where your eyes are looking: exploring the tendency to imitate another individual's gaze. *Neuroreport* 13, 2259–2264.

Ricciardelli, P., and Driver, J. (2008). Effects of head orientation on gaze perception: how positive congruency effects can be reversed. *Q. J. Exp. Psychol. (Hove)* 61, 491–504.

Salvucci, D. D., and Goldberg, J. H. (2000). "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the Eye Tracking Research and Applications Symposium* (New York: ACM Press), 71–78.

Sidner, C., Lee, C., Kidd, C., Lesh, N., and Rich, C. (2005). Explorations in engagement for humans and robots. *Artif. Intell.* 166, 140–164.

Staudte, M., and Crocker, M. (2009). "Visual attention in spoken HRI," in *HRI '09*, San Diego.

Tomasello, M. (2008). *Origins of Human Communication*. Boston: MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691; discussion 691–735.

van de Riet, W. A., Grezes, J., and de Gelder, B. (2009). Specific and common brain regions involved in the perception of faces and bodies and the representation of their emotional expressions. *Soc. Neurosci.* 4, 101–120.

Warneken, F., Chen, F., and Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Dev.* 77, 640–663.

Wätcher, A., and Biegler, L. T. (2006). On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Math. Program.* 106, 25–57.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 January 2012; paper pending published: 31 January 2012; accepted: 10 April 2012; published online: 03 May 2012.

Citation: Boucher J-D, Pattacini U, Lelong A, Bailly G, Elisei F, Fagel S, Dominey PF and Ventre-Dominey J (2012) I reach faster when I see you look: gaze effects in human–human and human–robot face-to-face cooperation. Front. Neurorobot. ***6****:3. doi: 10.3389/fnbot.2012.00003*