Facilitative Effects of Communicative Gaze and Speech in Human-Robot Cooperation

Jean-David Boucher, Jocelyne Ventre-Dominey, Peter Ford Dominey, Sascha Fagel, Gérard Bailly

Facilitative Effects of Communicative Gaze and Speech in Human-Robot Cooperation

Jean-David Boucher, Jocelyne Ventre-Dominey, Peter Ford Dominey
INSERM U846, 18 ave Doyen Lepine
69675 Bron Cedex, France
+33 472913475
Peter.dominey@inserm.fr

Sacha Fagel, Gerard Bailly
GIPSA-lab, UMR 5216 CNRS/Université de Grenoble
961 rue de la Houille Blanche – Domaine Univ. BP46
38402 Saint Martin d'Hères CEDEX. France
+ 33 476574711
Gerard.bailly@gipsa-lab.grenoble-inp.fr

ABSTRACT
Human interaction in natural environments relies on a variety of perceptual cues to guide and stabilize the interaction. Humanoid robots are becoming increasingly refined in their sensorimotor capabilities, and thus should be able to manipulate and exploit these communicative cues in cooperation with their human partners. In the current research we identify a set of principal communicative speech and gaze cues in human-human interaction, and then formalize and implement these cues in a humanoid robot. The objective of the work is to render the humanoid robot more human-like in its ability to communicate with humans. The first phase of this research, described here, is to provide the robot with a generative capability – that is to produce appropriate speech and gaze cues in the context of human-robot cooperation tasks. . 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.

Categories and Subject Descriptors
H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6) - User-centered design; Evaluation/methodology

General Terms
Human Factors

Keywords
Human-robot cooperative interaction, gaze, speech, pointing, deixis.

1. INTRODUCTION
Cooperation is one of the hallmarks of human cognition [1]. One of the central features of cooperation is the creation and manipulation of a shared plan between the two cooperating agents. This 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 [1]. Recent research has made progress in the development of a shared plan capability in the context of human-robot cooperation [2, 3, 4]. In this context the iCub robot has been provided with the capability to watch two humans perform a cooperative task and to use vision to detect actions, agents and goals, in order create and use a shared plan describing the cooperative action. That shared plan can then be used by the iCub to participate in achieving the shared goal, taking on either of the two possible roles [4]. There is an important limitation in this work however, which is related to the ongoing control of the cooperation by the use and monitoring of gaze.

The goal of the current research is to identify a basic set of gaze-related cooperative communication cues, from the literature and from our ongoing experiments, and to implement and test the efficacy of these cues on human performance in a cooperative interaction task with the iCub robot. We take a systematic approach in this endeavor, with a progressive introduction of communicative cues and evaluation of their impact. The current research thus reports on the effects of manipulation of goal-directed gaze cues on human performance in a cooperative task.

2. THE IMPORTANCE OF GAZE
One of the most central and important factors in the ongoing management of cooperative human interaction is the use of gaze to coordinate that one's interlocutor is present, paying attention, attending to the intended elements in the scene, checking back that all is ok [5]. In this context, gaze is highly communicative both in indicating one's proper attentional focus and in allowing the following of that of the interlocutor, and this importance is revealed in the specialization of brain systems dedicated to these functions [6].

Figure 1. Face-to-face human interaction laboratory set-up (GIPSA-Lab).

AFFINE'10, October 29, 2010, Firenze, Italy.

Thus while it is clear, even intuitively, that gaze is of central importance, one must ask what precisely are the characteristics of gaze control that will allow the most productive behavioral alignment between two partners in a cooperative task? In order to address this question, Fagel and colleagues have developed a human-human cooperation scenario in which speech and gaze are crucially involved [7]. The "Cube Game" task is illustrated in Figures 1 and 2. Two human subjects (the "informant" and the "manipulator") are seated face to face across a table. On the table is arranged a set of cubes that are labeled with the consonants {T, G, P, D, S, F, M, N, B}. Importantly, only the informant can see these consonant labels.

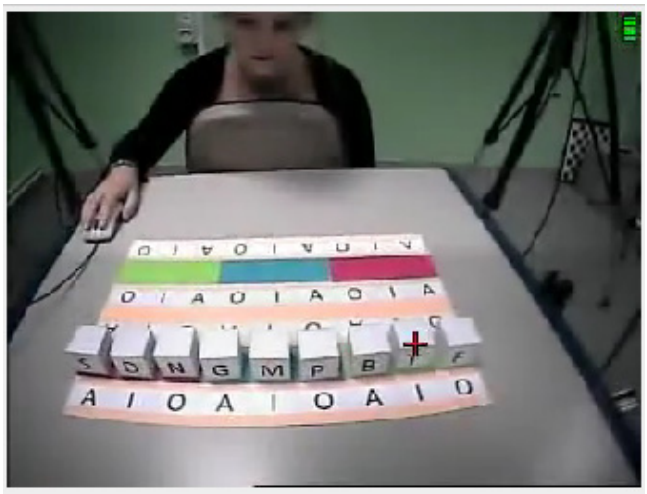

Figure 2. Point of view of the "informant" who can see the consonant labeled cubes (GIPSA-Lab).

The manipulator hears the consonant label of the target cube over a headset, and announces this to the informant. The informant (whose viewpoint is illustrated in Figure 2) searches for the cube, and then informs the manipulator of its (vowel, color) coordinates (which both informant and manipulator can see). The manipulator can then take the cube based on this specification. The game is cooperative by definition, because the informant needs the manipulator to announce the consonant label, and the manipulator needs the informant to find and announce the (vowel, color) label in order to finally get the target cube. The informant naturally scans the cube array while searching for the target consonant cube, and then holds her gaze there once the cube is found. In order to study the effects of the informant's gaze on the manipulator, two experimental conditions were studied: normal, and with the informant wearing sunglasses which render the eye-direction invisible to the manipulator.

Fagel & Bailly [8] recorded the search time (time required for the informant to locate and specify the location of the target cube), and the location time (time required for the manipulator to take the cube from the onset of the informant's command). They observed a significant effect of the glasses on the location time. This suggests that the manipulator uses the eye position information from the informant (who gazes at the target position she will announce) in order to anticipate that target position.

3. IMPLEMENTATION OF COOPERATIVE CUES IN A HUMANOID ROBOT

In this context, we set out to determine if such communicative effects of gaze could be used in the realm of human-robot cooperation. Specifically, we duplicate the cube game scenario, with the iCub robot playing the role of the informant. We can then directly manipulate the gaze of the iCub within the context of the task and determine if this has an impact on the performance of the naïve human manipulator.

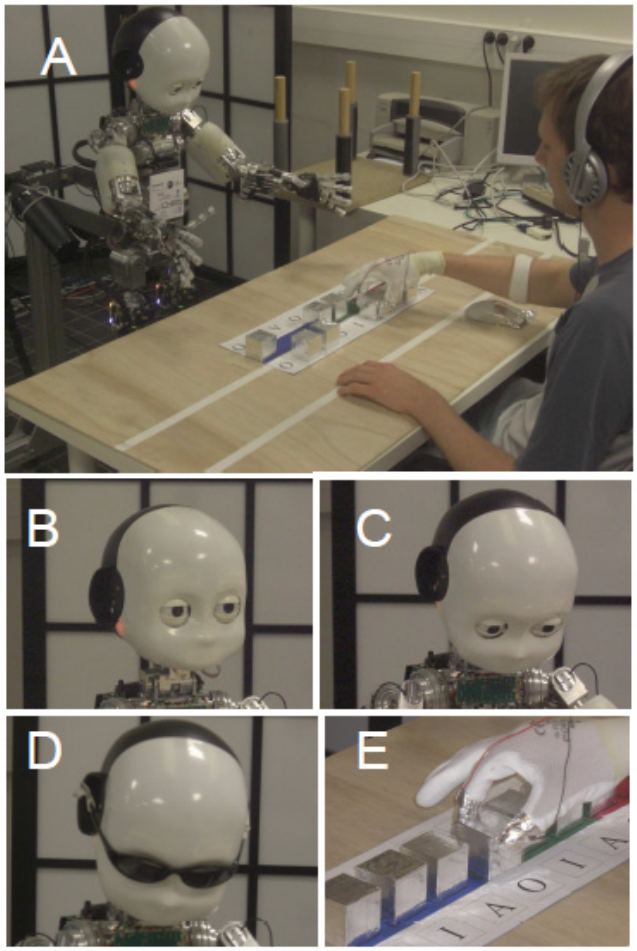

Figure 3. Human-Robot cooperation in the cube game. A. Cooperation set-up. B. iCub gaze when looking at the manipulator. C. Goal-directed gaze. D. Goal-directed gaze with sun-glasses. E. Example of manipulator taking a labeled cube.

3.1 The iCub Humanoid Platform

The iCub is a 53 degree of freedom humanoid robot with the body size of a 3-4 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.

3.2 The Cube Game with iCub

We followed the protocol of the cube game as developed by Bailly and colleagues [7,8]. In the game, the robot is always the informant. The robot is seated 50 cm from the aligned cubes on the playing surface (see Figure 3), with the manipulator (naïve human subject) sitting across the table. The manipulator 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. After a delay of 3 seconds (during which the manipulator repeats the cube consonant label), the robot looks away from the manipulator's face, and to the identified cube[1]. One second after initiating this combined eye-head gaze movement which has a total duration of 700ms, the robot then announces the location of the cube. The manipulator then reaches to the cube and moves it towards herself onto the labeled square.

3.3 Experimental Conditions

We studied the effects of robot (informant) communicative behavior on the naïve human subjects in thee conditions. In the Full-Gaze condition (Fig. 3C), the robot indicates the location of the target cube by a coordinated eye and head movement. The eye movement has a completion time of 200 ms, and the head movement a slower completion time of 700 ms. The eye and head are coordinated by an inverse kinematics solver that drives the motors in coordination. The saccade to the target thus terminates in the first 200 ms, then the eyes stay fixed on the target as the head completes its movement 500 ms later. In the "Sun-Glasses" condition (Fig. 3D), the human manipulator cannot see the iCub's eyes as the iCub is wearing sun glasses. Thus the initial positioning information provided by the eyes is not available. The manipulator must wait 700 ms for the completion of the head movement. In the head-fixed condition (Fig. 3B), there is no head nor eye movement generated by the iCub informant. Target cube position information is provided solely by speech. Note that an anticipatory motion of the hand off the start position and towards the target can be predicted in the Full Gaze and Sun Glasses conditions, as the deictic gaze motion precedes the speech command.

3.4 Performance Measures

In order to accurately measure the manipulator's 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 Figure 3 A and E). One set of contacts is connected to metallic contacts on the index and thumb of a glove that the manipulator wears on her right hand. When these contacts come into physical contact with a metallic cube, the mouse registers a right down click. 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. To initiate a trial, the subject puts her palm on the resting position marker and this is recorded as a right down click. As the trial proceeds, when the subject raises her hand from the starting position this is recorded as the movement onset time. Contact with the cube is recorded to generate the movement completion time. Finally grasping time is the duration from the movement onset to movement completion. Five naïve subjects were exposed to 10 games in each of three conditions for a total of 30 games each, with the total duration of approximately 60 minutes, divided into two 30 min sessions. The three conditions were full gaze, sun-glasses, and head fixed. The first two conditions correspond to those used by Fagel & Bailly [8]. The third condition is difficult or impossible to be realized by human subjects, who cannot inhibit their natural gaze towards the target cubes. The use of the iCub allows us to explore this condition.

[1] In [8] the manipulator must hear the informant announce the consonant label. Here the iCub "knows" the label in advance but waits for the processing time as if attending to the informant.

4. RESULTS.

The results are presented in the context of the three experimental conditions, full gaze, sun-glasses gaze and head fixed. The principal results can be visualized in Figure 4.

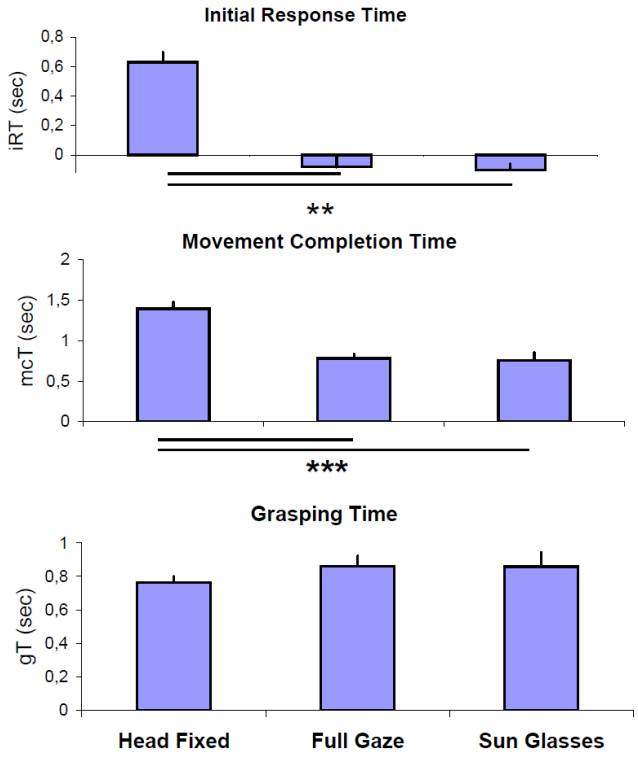

Figure 4. Summary behavior for 5 naive subjects, in the three Head Fixed, Full Gaze, and Sun Glasses conditions. Results in seconds. Initial response (delay from robot cube specification until human lift off from starting position). Movement completion time (delay from robot specification until grasp of target cube). Grasping time (delay from human lift off until grasp).

We illustrate three performance measures: the initial response time, the movement completion time, and the grasping time. Recall that the initial movement time is characterized as the numerical difference in seconds between the initiation of the speech signal from the robot informant, and the lifting of the

human manipulator's hand from the starting position. It should be made clear that the robot's gaze can aid the manipulator to identify the target color and precise location (the A, I or O specification) before the target has been specified by speech. In the limit case, the human manipulator can actually anticipate the speech signal and place her hand near the colored target location before the final target has been specified. This can be observed in the full gaze and sun glasses conditions.

The movement completion time is characterized as the delay between the onset of the informant (robot) specification of the target (vowel, color) location, and the human subject's first touch of the corresponding cube.

As illustrated in Figure 4 the three conditions (Head Fixed, Sun Glasses and Full Gaze) had visible influence on both the initial response time and movement completion time. This was confirmed by repeated measures ANOVA. For the initial response time the main effect of condition was significant $F(2,8) = 23$, $p<0.001$. Initial response times were significantly different for head fixed vs. sun glasses and full gaze $p = 0.0011$ and $p = 0.0014$ respectively (Sheffe post hoc).

For the movement completion time the main effect of condition was significant $F(2,8) = 41$, $p<0.001$. Movement completion times were significantly different for head fixed vs. sun glasses and full gaze $p < 0.001$ and $p < 0.001$ respectively (Sheffe post hoc).

For the grasping time the main effect of condition was not significant $F(2,8) = 1$, $p = 0.39$.

5. DISCUSSION

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 [e.g. 9]. We extend this in a complimentary 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 our experimental protocol, the robot's speech specification of the target cube is slightly preceded by it's gaze to that target, and we could thus predict that that subjects will attend to this gaze information and begin to move their hand towards the target before the speech signal, exploiting the information in the gaze signal. This corresponds to the negative (anticipatory) values for the initial response times in Figure 4, for the conditions in which gaze was present (with or without shielding of the eyes by the sunglasses). This advantage for the full gaze and sun glasses conditions was likewise transferred to the movement completion times. The essentially sensorimotor aspects of the grasping motion itself were not influenced by our experimental manipulation of gaze. This indicates that independent of the condition, once a subject lifted their hand from the starting position, the time to complete the movement was not effected by the gaze conditions.

Fagel & Bailly [8] observe a significant "sun glasses" effect, i.e. when the manipulator could not see the informant's eyes, task completion times were increased. For the iCub we did not observe this effect. Interestingly, Fagel & Bailly [8] indicate, that for their subjects, while the eyes fixated the specific target, the head position was not specific for each target location, but rather indicated the approximate region. Thus the sun glasses deprived subjects of partial information. In contrast, the iCub head position was directed precisely at each target, and was thus redundant with eye position at the end of the movement. This allows us in the future to test the prediction that if the iCub head movements are rendered less precise, indicating the region but not final position (as the human's) then the sunglasses will have an effect. This raises the more general issue that human gaze is highly dynamic, moving from the shared target of attention to the face and eyes of the interlocutor and back with an elevated frequency. We will begin to characterize these dynamics in more detail and asses their influence on the human capacity to cooperate with robots with increasing communicative fidelity.

6. ACKNOWLEDGMENTS

This research has been supported by the French ANR Project Amorces.

7. REFERENCES

[1] (2005) Understanding and sharing intentions: The origins of cultural cognition, Beh. Brain Sc;. 28; 675-735.

[2] Dominey , P.F., Warneken, F. (2009) The origin of shared intentions in human-robot cooperation, New Issues in Psychology

[3] Lallee S, Dominey PF & Warneken F (2009) Learning To Cooperate by Observation, Epigenetic Robotics, Venice.

[4] Lallee S; Madden C, Hoen M, Dominey PF (2010) Linking Language with Embodied and Teleological Representations of Action for Humanoid Cognition, in press, Frontiers in NeuroRobotics

[5] Kendon, A. (1967). "Some functions of gaze-direction in social interaction." Acta Psychologica 26: 22-63

[6] Pourtois, G., D. Sander, M. Andres, D. Grandjean, L. Revéret, E. Olivier and P. Vuilleumier (2004). "Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals." European Journal of Neuroscience 20: 3507-3515.

[7] Bailly, G., S. Raidt, et al. (2010). "Gaze, conversational agents and face-to-face communication." Speech Communication - special issue on Speech and Face-to-Face Communication **52**(3): 598–612.

[8] Fagel, S. and G. Bailly (submitted). On the importance of eye gaze in a face-to-face collaborative task. Workshop on Affective Interaction in Natural Environments (AFFINE), Firenze, Italy.

[9] Sidner, C. L., C. Lee, Kidd CD, Lesh N, Rich C (2005). Explorations in engagement for humans and robots, Artificial Intelligence 166 (2005) 140–164.Tomasello M, Carpenter M, Call J, Behne T, Moll HY