Beaming the Gaze of a Humanoid Robot

Gérard Bailly, Frédéric Elisei, Miquel Sauze

capture system drive back Nina's head and eye movements: the HMD actually embeds an Arrington® binocular eye tracker and the head movement is tracked thanks to 5 reflexive markers glued to the helmet. Using Nina's PID controllers, the latency between tracked and executed movements is around 60ms. Saccades of the robot's eyes are triggered by estimating the fixation point from the binocular view and centering this point via a mapping that associates fixation points to eye angles.

Figure 1. Our beaming platform consists of a HMD with an embedded Arrington® binocular eyetracker. The HMD receives input from Nina's eye cameras and ear microphones. Movements of the HMD are tracked by a Qualisys® motion capture (MoCap) system. The eyetracker and the MoCap system drive back Nina's head and eye movements.

4   A FIRST BEAMING EXPERIMENT

We explore the impact of a continuous monitoring of the robot's gaze (eyes and head movements) by an experienced pilot on a joint task. We already have shown [5] that humans may benefit from the robot's visual attention when jointly scrutinizing a scene.

Interaction scenario. The interaction scenario involves a trained pilot and one naïve manipulator. The manipulator receives confidential oral instructions (via earphones) from the system, each consisting of moving one cube from its current place to another random place. The trick is that the system refers to the cube by an identifier (a letter) that is only visible to the pilot (see Figure 2): the manipulator has thus to ask the pilot about the current location of the cube before moving it. The next instruction is given by the system two seconds after the cube has been put in the right place. A game consists of moving as many cubes as possible within three minutes while avoiding mistakes. Each manipulator plays successive games under three different beaming conditions (within participants design), performed in counterbalanced order:

- *Fixed*: The robot remains with a fixed leaned head and eye orientation, allowing the pilot to see the entire chessboard
- *Head*: The robot's head follows the pilot's head movement, but robot eyes orientation is fixed
- *Head+Eyes*: The robot's head & eyes follow the pilot's head & eyes movements

The pilot's speech and eye blinks are always transmitted.

By varying the conditions of visibility of the pilot's movements, we test here their impact on the speed, robustness and comfort of the joint task. In fact, the task could be performed entirely by oral dialog, but we expect that head and eye movements will ease interaction and will speed turns. At game onset, 7 cubes are randomly distributed on the chessboard, one per column: we thus have a spare column. Via earphones, the system instructs the manipulator to move a cube to one random tile of the spare column (e.g. "move N to B3" in Figure 2). The manipulator then asks the pilot about

---

Beaming the Gaze of a Humanoid Robot

Gérard Bailly   Frédéric Elisei   Miquel Sauze

GIPSA-Lab, Université Grenoble Alpes/CNRS
11, rue des Mathématiques, 38401 St Martin d'Hères - France
firstname.lastname@gipsa-lab.fr

ABSTRACT

We here propose to use immersive teleoperation of a humanoid robot by a human pilot for artificially providing the robot with social skills. This so-called *beaming* approach of learning by demonstration (the robot passively experience social behaviors that can be further modeled and used for autonomous control) offers a unique way to study embodied cognition, i.e. a human cognition driving a controllable robotic body.

Theme: Enabling Methods
Categories and Subject descriptors: H.4 [**Information Systems Applications**]: Miscellaneous
Keywords: human-robot interaction; immersive teleoperation.

1   INTRODUCTION

We describe an immersive teleoperation system that allows a human pilot to remotely interact with a human partner via a co-located local robotic incarnation. Our system enables the pilot to control the movements of the head and eyes of the robot with its own movements, while perceiving the local scene in which the robot is immerged through the robot's eyes and ear microphones.

2   NINA: A TALKING & GAZING ICUB

The standard iCub robot [1] has 53 actuated degrees of freedom including 3 angles for the elevation, azimuth and tilt of the neck and 3 angles for the elevation, azimuth and vergence of the eyes that embed Dragonfly cameras. The inner head houses an on-board PC104 controller, the electronic boards to control the electric motors, a gyro sensor and two camera boards. The eyes embed stereo cameras in a swivel mounting with degrees-of-freedom similar to human eyes (azimuth, elevation and vergence).

IIT and GIPSA recently redesigned the iCub head [2] in order to enhance its communicative skills (see Figure 1). The new robot, called Nina, benefits from a new implantation of the eye tilt mechanism – together with the suppression of lower eyelids – that freed the space for a Visaton BF32 loudspeaker, two high-quality ear microphones plugged into the auricles and five additional motors for articulating the jaw and the lips. The mechatronic articulated head is also covered with a stretch fabric.

3   BEAMING NINA

The concept of "beaming" [3, 4] is intrinsically asymmetrical: a *beaming* platform gives the pilot the strong sensation of ownership regarding his/her body representation at destination, while local users should experience a strong sense of physical presence and situation awareness from the avatar.

Our current *beaming* platform is shown in Figure 1. The pilot is equipped with a head-mounted display (HMD) to remotely perceive the local scene and motion-capture devices to act reciprocally on it. The HMD receives stereo inputs from Nina's eye cameras and ear microphones. An eyetracker and a Qualisys® motion

the position of the cube by saying "cube N?". The pilot searches for the cube and instructs back the manipulator about its position (saying "column H"). The manipulator then grasps and moves the cube to the requested position (i.e. from H8 to B3) while uttering the cube's destination (i.e. "to B3"). Interacting language is French. We report here experiments performed by one unique experienced male pilot interacting with ten naive subjects.

Figure 2: A sample of the pilot's view on the chessboard via the right-eye robot (640x480 Dragonfly camera). The system instructs the manipulator facing the robot to a cube whose identifier (one letter) is only visible by the pilot via the robot's eyes. The other faces of the cubes are left blank.

Performance measures and questionnaires. Performance is characterized by the number of cubes correctly placed in each game. Manipulators rated the subjective *mean opinion scores* (*MOS*) of conditions. They rated the quality of their experience about the usefulness of nonverbal behaviors (USE), the robot's visual attention to the chessboard (VAC) and to them (VAS).

Subjective ratings. Although some subjects did report not to pay attention to the robot, subjective appreciations of the subjects evolve as expected and do not depend on order of experienced beaming conditions (see Figure 3). This ranking is highly significant: the *Head+Eyes* condition receives the highest ranking for all judgments ($p<10^{-6}$) while the *Head* condition stands in the intermediate position ($p<0.004$). These judgments should be taken with extreme care: they rate the pilot's embodiment for that precise interactive task with these robot/pilot pairs. They however stress the importance of redundancy and complementarity of multimodal behaviors: gaze is a complex by-product of head and eye movements and interlocutors are surely disturbed by the fixed eye direction that does not anticipate/drive the head motion.

In contrast, these positive judgments have no significant impact on performance, i.e. number of moves per minute. Possible explanations are numerous: the full immersive control of the robot slows down the pilot, the decoding of the current multimodal behaviors increases the subjects' cognitive load, etc. Another possible problem is the current absence of jaw and lips movements that should accompany speech production and likely attract user's attention out of the table. We effectively plan to work this out.

Pilot's behavior. The number of degrees of freedom (DOF) effectively controlled by the pilot's movements has an important effect on the behavior of the trained pilot. Although the pilot is free to move his own eyes and head, the audiovisual feedback he gets from his actions strongly impacts his motor plans. Figure 4 displays the distributions of head movements and the number of fixations/s according to the three beaming conditions. An analysis of variance shows that the beaming conditions have significant effects on variances of all cues: *Head* exhibits the largest variance for head movements and lowest fixation rate, *Head+Eyes* distributes variance among all DOF while *Fixed* almost freezes all head DOF. These behavioral profiles demonstrate implicitly that the immersive teleoperation of the iCub robot by the trained pilot is

effective and that the pilot takes into account the affordances of the robot when driving his own body. Besides being an elegant way to provide robots with social skills, the beaming technique is thus a very powerful technique to study the recalibration of motor control in interaction when perturbed, as well as studying how human cognition makes the best – executive as well as social – use of available body kinematics.
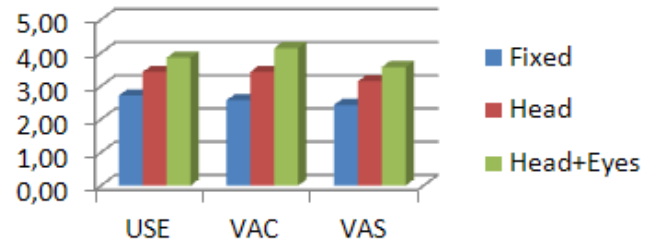
Figure 3: Mean MOS judgments for the three conditions: usefulness of nonverbal behaviors (USE), rating of the robot's visual attention to the chessboard (VAC) and to them (VAS).
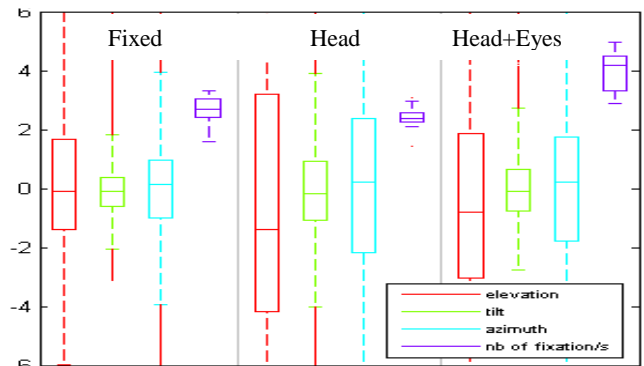
Figure 4: Distributions of the 3 head angles (°) and the number of fixations per seconds (cumulated for all 10 subjects) observed for the three conditions.

CONCLUSIONS

We describe an original immersive teleoperation that enables a pilot to offer situation-aware gaze skills to our robot. Our subjects rated favorably mobile head and eyes, while the trained pilot implicitly takes into account and reacts to the modifications of the robot's affordances. Beaming provides an elegant way for learning multimodal behavioral models by active demonstration, while both solving the challenge of scaling human behaviors to HRI and studying social acceptance of faked cognitive robots, until these are endowed with autonomous social skills.

REFERENCES

[1] Parmiggiani, A., et al., *The design of the iCub humanoid robot.* International Journal of Humanoid Robotics, 2012. 9(4): 24 pages.

[2] Parmiggiani, A., et al. *An articulated talking face for the iCub.* in *Humanoids.* 2014. Madrid.

[3] Steed, A., et al., *Beaming: an asymmetric telepresence system.* IEEE Computer Graphics and Applications, 2012. 32(6): p. 10-17.

[4] Slater, M., et al., *First person experience of body transfer in virtual reality.* PLoS ONE, 2010. 5(5): e10564.

[5] Boucher, J.-D., et al. (2012) *I reach faster when I see you look: Gaze effects in human-human and human-robot face-to-face cooperation.* Frontiers in neurorobotics, 6 pages DOI: 10.3389/fnbot.2012.00003.