

# Learning robotic eye–arm–hand coordination from human demonstration: a coupled dynamical systems approach

Luka Lukic · José Santos-Victor · Aude Billard

Received: 26 December 2012 / Accepted: 3 February 2014 / Published online: 26 February 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** We investigate the role of obstacle avoidance in visually guided reaching and grasping movements. We report on a human study in which subjects performed prehensile motion with obstacle avoidance where the position of the obstacle was systematically varied across trials. These experiments suggest that reaching with obstacle avoidance is organized in a sequential manner, where the obstacle acts as an intermediary target. Furthermore, we demonstrate that the notion of workspace travelled by the hand is embedded explicitly in a forward planning scheme, which is actively involved in detecting obstacles on the way when performing reaching. We find that the gaze proactively coordinates the pattern of eye–arm motion during obstacle avoidance. This study provides also a quantitative assessment of the coupling between the eye–arm–hand motion. We show that the coupling follows regular phase dependencies and is unaltered during obstacle avoidance. These observations provide a basis for the design of a computational model. Our controller extends the coupled dynamical systems framework and provides fast and synchronous control of the eyes, the arm and the hand within a single and compact framework, mimicking similar control system found in humans. We validate our model for visuomotor control of a humanoid robot.

**Keywords** Visuomotor coordination in reach and grasp movements · Coupled dynamical systems · Forward planning · Obstacle avoidance

## 1 Introduction

Manipulation and grasping skills are complex and rely on the conjunction of multiple modalities, including vision, tactile and proprioceptive information (Prablanc et al. 1979; Jeannerod 1984; Purdy et al. 1999). Vision provides important information in the early stages of motion planning (Prablanc et al. 1979; Abrams et al. 1990; Spijkers and Lochner 1994; Rossetti et al. 1994). It is also used to perform close-loop control to drive the hand in space unobstructed visually (Abrams et al. 1990; Paulignan et al. 1991), while tactile information becomes crucial in the last stage of prehension and to compensate when vision cannot be used<sup>1</sup> (Jeannerod 1984; Purdy et al. 1999). Vision is particularly useful to plan motion so as to avoid obstacles without touching them (Johansson et al. 2001). It also enables to react rapidly in the face of a sudden perturbation, such as an obstacle entering the workspace (Aivar et al. 2008). There is a tight coupling between visual and motor modalities when driving prehensile motion (Prablanc et al. 1979; Land et al. 1999; Johansson et al. 2001). While this coupling has been documented at length in the literature in free space motion (Johansson et al. 2001; Hayhoe et al. 2003; Bowman et al. 2009), little is known about how this coupling is exploited to enable fast and reliable obstacle avoidance, and in particular when the obstacle appears after the onset of motion. Such fast and online control of hand motion in response to visual detection of obstacle

---

L. Lukic (✉) · A. Billard  
Learning Algorithms and Systems Laboratory, Ecole  
Polytechnique Fédérale de Lausanne, EPFL-STI-I2S-LASA,  
Station 9, 1015 Lausanne, Switzerland  
e-mail: luka.lukic@epfl.ch

A. Billard  
e-mail: aude.billard@epfl.ch

J. Santos-Victor  
Computer Vision and Robotic Laboratory,  
Instituto Superior Técnico, Avenida Rovisco Pais 1,  
1049-001 Lisbon, Portugal  
e-mail: jasv@isr.ist.utl.pt

<sup>1</sup> Humans can perform prehensile actions without visual feedback, by relying on tactile and acoustic senses.

is crucial for humans, but also for robots. Indeed, in spite of impressive advances in robotics over the last decades, robots are still far from matching human's versatility in the control of their motion, even when performing the most simple reach and grasp motion.

This paper aims at developing a computational model of the visuomotor coupling between the eye–arm–hand systems, that explains how this coupling is modulated by the presence of an obstacle. We conduct a motion study of reach and grasp motion in the presence of obstacle. The human study provides quantifiable information about the eye–arm–hand coupling to support the design of the model's parameters. To demonstrate the feasibility of using this mechanism for robot control, we implement our model to control visually guided prehensile motion in the iCub humanoid robot.

We next provide a short review on existing works, focusing on the role of visual information in guiding manipulation, visuomotor coordination mechanisms in humans and the state of the art in robotic visually aided manipulation, and we conclude this section with an overview of our contributions.

### 1.1 The role of visual information in guiding manipulation

Vision provides a plethora of by far the most valuable and most reliable information about the state of the environment on which the planning and motor systems depend heavily. The object's extrinsic properties (spatial location and orientation) are used to control the reach component, whereas the object's intrinsic properties (shape, size, weight, centroid and mass distribution) are used in programming the grasp component (Jeannerod 1984). The role of vision in manipulation is best shown in manipulation experiments where visual feedback is deprived by modulating experimental conditions.

Several studies have shown that manipulation without any visual feedback in highly structured, static scenarios can almost match the performances of the full-vision manipulation (Castiello et al. 1983; Purdy et al. 1999). After a number of practice trials, manipulation of subjects who did not have any visual feedback only slightly differed from full-vision manipulation in terms of the kinematic measures of both the reach and grasp components. However, if manipulation without visual feedback is performed in an unstructured environment, without previous kinesthetic assistance from a teacher or extensive trial-and-error learning, the performances (e.g., overall success rate, accuracy of reaching and speed of movement) drastically degrade compared with trials where vision was not deprived (Purdy et al. 1999).

Vision is used to guide every stage of prehensile movements, from pre-planning, initial reach, high-speed mid-section of the movement, to the deceleration and grasping

phases. Prablanc et al. (1979) and Rossetti et al. (1994) showed that seeing the limb before the onset of the movements improves the reaching accuracy. In addition to this, Pelisson et al. (1986) found that the initial information about the target affects the final reaching accuracy. Similarly, the sight at the current position of the limb and the movement goal in the later stage of the movements improves the end point accuracy (Prablanc et al. 1979; Pelisson et al. 1986). In studies of manipulation where no visual feedback on the moving limb (Gentilucci et al. 1994; Berthier et al. 1996) and on the target (Jakobson and Goodale 1991) is available, a dramatic increase in the overall movement time and the grip aperture was observed. Finally, visual information assists fine control of the arm and hand in the closing phase of grasping (Paillard 1982). The gaze is driven to the grasping points on the target object during a prehensile task, for the purpose of planning reliable placement of the fingers (Brouwer et al. 2009). These studies suggest that vision is used for online control of both the reaching and grasping components of a prehensile movement.

A number of studies have shown that both peripheral and foveal vision contribute to reaching and grasping. Sivak and MacKenzie (1990) found that when central vision was blocked, it affected both the transport and grasp components (longer movement times, lower peak accelerations and peak velocities, larger maximum grip apertures and longer time after the maximum grip aperture). When peripheral vision was not available, however, they observed that it affected the transport component only, and the grasp component remains unaltered. In their follow-up study, González-Alvarez et al. (2007) found that peripheral and foveal visual cues jointly contribute to both reaching and grasping.

Further evidence that vision is used for online control of movements comes from perturbation studies. In the study by Paulignan et al. (1991), subjects were able to instantly modulate, by relying on visual feedback, the arm and hand movements with respect to online perturbations of the position and shape of the target object, with only minimal increase in the response time ( $\sim 100$  ms) compared with motion in the absence of perturbations. Aivar et al. (2008) studied adjustments of the hand movements with respect to abrupt online perturbations of obstacles and/or the target. They found similar latencies to those reported by Paulignan et al. (1991) for the responses to the perturbations of the target position and slightly longer adaptation latencies for the obstacles.

### 1.2 Human visuomotor coordination

Human visual and motor systems are not independent, and they operate in coordination and share control signals adapting to mutual demands, even when doing simple and well-practiced routines (Land et al. 1999; Hayhoe et al. 2003). A body of literature documented how the gaze precedes

motion. The gaze shows an anticipatory strategy leading a whole-body movement during navigation (Grasso et al. 1998; Hicheur and Berthoz 2005; Rothkopf and Ballard 2009). The gaze precedes the arm and the hand movement in manipulation tasks with a tool in the hand (Johansson et al. 2001). Similar pattern, the gaze leading the arm, is observed in a task where subjects contacted multiple target objects arranged in a sequence (Bowman et al. 2009). Abrams et al. (1990) found that the gaze leads limb movements in rapid tasks as well. Furthermore, it is also observed that gaze leads the arm and the whole-body movements in reach-for-grasp tasks (Land et al. 1999; Hayhoe et al. 2003; Hesse and Deubel 2011). Physiological studies of prehensile motion report that the arm transport and the hand preshape components are coordinated by the motor system in reach-for-grasp maneuvers, even in the presence of perturbations (Caciello et al. 1993; Haggard and Wing 1995). Furthermore, there is a strong evidence that control signals also flow from the hand to the eyes, not only in the opposite direction (Fisk and Goodale 1985; Neggens and Bekkering 2000).

In addition to the physiological studies that measure the visible manifestations of visuomotor coupling, neuroscience studies in primates show evidence of a strong correlation in the neural activity displayed in the areas devoted to vision, motor control, attention and gaze movements. More specifically, joint neural activation, as well as tight structural connectivity, has been observed between the posterior parietal cortex, which is involved in transforming visual signals into motor plans, with the frontal planning and motor areas and the inferior parietal cortex involved in preparation of saccadic eye movements and attention shifts (Rizzolatti et al. 1997; Andersen and Cui 2009; Baldauf and Deubel 2010). This coupling is very important in terms of information processing. If the systems are not properly synchronized, valuable information is lost. Additionally, an explicit synchronization aids a fast, reactive behavior, because control signals are directly transmitted between different control modules, without unnecessary delays.

While we have emphasized until now the importance of active gaze control to drive arm–hand motion, it is noteworthy that humans can also grasp an object without fixating it and even perform more complicated tasks such as obstacle avoidance by solely relying on peripheral vision (Prablanc et al. 1979; Abrams et al. 1990; Johansson et al. 2001). In spite of the fact that humans may reach without looking at the target, in natural and unrestricted tasks, the gaze seems to lead the arm–hand movement. This mechanism is likely a safeguard mechanism to ensure accurate reaching in the face of obstacles. Indeed, when saccades to the target and obstacle were prohibited, significantly decreased manipulation accuracy was observed (Abrams et al. 1990; Johansson et al. 2001) and manipulation resulted in frequent collisions with the obstacle (Johansson et al. 2001). These experiments pro-

vide a further evidence that coupling between active vision and the motor system is an important and fundamental mechanism, synchronously orchestrated between different regions in the central nervous system (CNS).

### 1.3 Robotic visually aided manipulation and obstacle avoidance

Identifying and modeling the mechanisms at the basis of human visuomotor control in the presence of obstacle provide a promising research direction to improve the design of similar controllers in robots. In our work, we exploit three paradigms (and the interplay between them) from the human visuomotor system can endow robots with a higher degree of dexterity and autonomy: *active vision* that is *coupled and synchronized* with the motor system constituting a coherent, but still modular, mechanism, which can *rapidly react to perturbations* in the environment. Some computer vision problems that are inherently ill-posed when using passive vision become well-posed when employing an active vision strategy<sup>2</sup> (Gibson 1950; Bajcsy 1988; Bajcsy and Campos 1992). Aloimonos et al. (1988) and Ballard (1991) have shown that an observer engaged in the active vision strategy gains a number of advantages over a passive observer, namely in terms of the cost of visual computation, the stability of algorithms and the uniqueness of solutions when determining shapes, determining structure from motion and computing depth. In active visual systems, visual servo control is computationally easier and more robust to errors in measurements as well (Ballard 1991). Coupling mechanisms between different control modules play an important role for ensuring a proper coordinated execution of complex tasks, such as visually guided reaching where the torso, head (including the eyes), arm and hand are simultaneously engaged. A proper coordination pattern between modules is especially crucial when performing prehensile tasks in the face of perturbations (Shukla and Billard 2011). Finally, a real-world environment can be rather highly dynamic and unpredictable. An agent must be able to re-plan and react in a time range of several milliseconds to changes that can happen unexpectedly. Not being able to rapidly and synchronously react to perturbations can cause fatal consequences for both the robot and its environment.

Solutions to robotic visual-based reaching follow either of two well-established approaches: techniques that learn visuomotor transformations (Hoffmann et al. 2005; Natale et al. 2005, 2007; Hulse et al. 2009; Jamone et al. 2012), which operate in an open-loop manner, or visual servoing

<sup>2</sup> Active vision systems employ gaze control mechanisms to actively position the camera coordinate system in order to manipulate the visual constraints.

techniques (Espiau et al. 1992; Mansard et al. 2006; Natale et al. 2007; Chaumette and Hutchinson 2008; Jamone et al. 2012), which are closed-loop methods. Techniques that learn the visuomotor maps are very appealing because of their simplicity and practical applications. However, these methods suffer from several drawbacks. The models of the visuomotor transformations are learned by using exploratory schemes employed by a robot that are similar to babbling employed during infant development (Vernon et al. 2010). The number of exploratory movements that the robot needs to visit during the exploration is usually of the order of several thousands, or even higher. Such extensive exploration, needed to learn a model, limits the applicability of these methods because it is highly inefficient in time and energy spent. The accuracy of the reaching movement is limited by the accuracy of the eye–arm mapping estimate. Moreover, the learned mapping profile cannot be modulated. Finally, the reaching path is generated by relying on interpolation between the starting arm state and the computed goal arm state. On the other hand, visual servoing approaches control the speed of the arm, based on measurements of the visual error between the hand and the target. This approach ensures zero-error reaching, but it requires having the target object and the hand simultaneously in the field of view. Visual servoing does not allow us to produce a family of human-like motion profiles in reaching tasks. The previous work done on the visuomotor coordination did not address the synchronization pattern of the arm transport and grip component. A control policy of a robotic hand (or a gripper) is usually a pre-programmed routine that is invoked after the arm reaches the target object, thus its control mechanism is not embodied in the eye–arm control, as in humans.

Robots operating in cluttered environments have to be able to plan their motion avoiding collisions with objects in the workspace. There is a large number of obstacle avoidance methods, and providing a broad review is not our intended goal. We now provide a brief synopsis of the main trend across these approaches. Recently, the most popular methods are sampling-based algorithms (Kavraki et al. 1996; Kuffner and LaValle 2000). Sampling-based algorithms are very powerful, but cannot meet the demands of rapid motion planning that humans perform almost effortlessly in a fraction of a second. Additionally, robotic obstacle avoidance methods do not consider how gaze control is involved in the process of obtaining information about the state of obstacles and targets, and they usually assume that environment is somehow known beforehand. Seara et al. (2003) developed an algorithm to actively control the gaze of a humanoid robot in order to support visually guided walking with obstacle avoidance. However, in robotic obstacle avoidance applications, involving manipulation information about the environment is obtained either by using passive stereo systems (Khansari-Zadeh and Billard 2012), or by relying on some special sensors such

as Microsoft Kinect<sup>TM</sup> and laser rangefinders<sup>3</sup> (Srinivasa et al. 2012). Having a gaze control strategy in obstacle avoidance is crucial in order to fixate obstacles. Fixations at the obstacles provide accurate visual information about their state, and these information are used to proactively guide the arm–hand system. Failure to provide visual information about obstacles can result in fatal collisions.

#### 1.4 Our contribution

This paper proposes a novel computational model of the visuomotor control when performing reaching and grasping motion in the presence of obstacles. To guide our modeling, we conduct a human study in which 8 volunteers perform reach and grasp motion to a single target in the presence of an obstacle. We analyze the kinematics of the eye, arm and finger motion to provide quantitative measurements on the phase relationships across these limbs. We extend the coupled dynamical systems (CDSs) framework, originally used for arm–hand coordination (Shukla and Billard 2011), to model the eye–arm–hand coordinated pattern measured in the human study.

The parameters of our model are estimated based on the data recorded in the human study. Our approach contributes to a better understanding of visually guided reach and grasp motion in humans. Furthermore, it provides a novel approach to generate close-loop visuomotor servoing in robot control.

We extend our CDS framework for visuomotor coordination to encapsulate: (a) model of the eye–arm–hand coupling and (b) modulation by an obstacle. In our work, we exploit a biologically inspired notion of forward models in motor control (Wolpert et al. 1998, 2001) and use a model of the dynamics of the reaching motion to predict collisions with objects in the workspace when reaching for the target object. We use the observation from the human study that the obstacle may act as an intermediary target, in order to develop our obstacle avoidance scheme. The objects, which are tagged as obstacles after propagating the forward model, are treated as intermediary targets for the visuomotor system. This approach results in a simple and computationally lightweight scheme for obstacle avoidance. As an alternative to computationally costly sampling-based algorithms (Kavraki et al. 1996; Kuffner and LaValle 2000), our approach uses the ability of dynamical systems (DSs) to instantly re-plan motion with the presence of perturbations. In our obstacle avoidance scheme, the gaze is an important element of the coupled visuomotor mechanism that is actively controlled and tightly bound to manipulation requirements and plans. We demonstrate the usefulness of this model for robot control, by implementing it in experiments on real-time obstacle

<sup>3</sup> These sensors are not controlled in terms of the active vision paradigm.

avoidance in simulation and on the real humanoid robot, iCub (Metta et al. 2010).

The work presented here was published in a preliminary form in Lukic et al. (2012). The present paper extends our previous work in five ways: (a) it provides a more detailed description of the computational model and system architecture, (b) it provides a more comprehensive literature review, (c) the human experiment is done with more subjects and with a more precise motion capture system, (d) we substantially extend the analysis of the human study, and (e) it presents more robot experiments, verifying the presented approach.

The rest of the paper is structured as follows. Section 2 describes the experimental procedure with human subjects and analysis of the recorded data. Section 3 explains our approach to eye–arm–hand coordination, planning and obstacle handling. In Sect. 4, we present results of our robotic experiments. Finally, Sect. 5 is devoted to a general discussion.

## 2 Human motion study of reaching and grasping with obstacle avoidance

We start from the hypothesis that the eyes precede arm motion, so as to guide the planning of the arm transport component. There is ample evidence of such saccadic eye motion toward the target during reaching; see, e.g., (Land et al. 1999; Johansson et al. 2001; Hayhoe et al. 2003; Hayhoe and Ballard 2005); however, few studies have analyzed visuomotor behavior in trials where the position of the obstacle was systematically varied. We assume that the obstacle acts as an intermediary target when performing obstacle avoidance. This movement-segmented strategy substantially reduces the complexity of motor control compared with the holistic control policy (Alberts et al. 2002; Johansson et al. 2009; Hesse and Deubel 2010). Furthermore, we hypothesize that there exists a visuomotor forward control scheme in which the presence of the obstacle is used to modulate the path of the arm. This modulation depends on the distance of the original path to the target. We also assume that the obstacle avoidance maneuver consists in passing the obstacle on the side of the obstacle where the collision would have occurred. This choice participates in a minimum effort strategy with only a small modulation of the intended path. We report our analysis of the visuomotor obstacle avoidance scheme in the following sub-sections. Figure 2 shows snapshots taken from the WearCam video illustrating the mechanism of the gaze leading arm motion and fixating the obstacle on the path when reaching the target.

The first part of this section describes the experimental procedure followed during our human motion study. In the second part, we analyze the results of this study and state our

findings of visuomotor coordination that constitute a basis for developing our computational model.

### 2.1 Subjects

Eight unpaid subjects from the university staff participated in this experiment (5 males and 3 females; mean age 27.1 years and SD 3 years). Subjects were right-handed and did not have any neurological or ophthalmological abnormalities. Subjects were unaware of the purpose of the experiment.

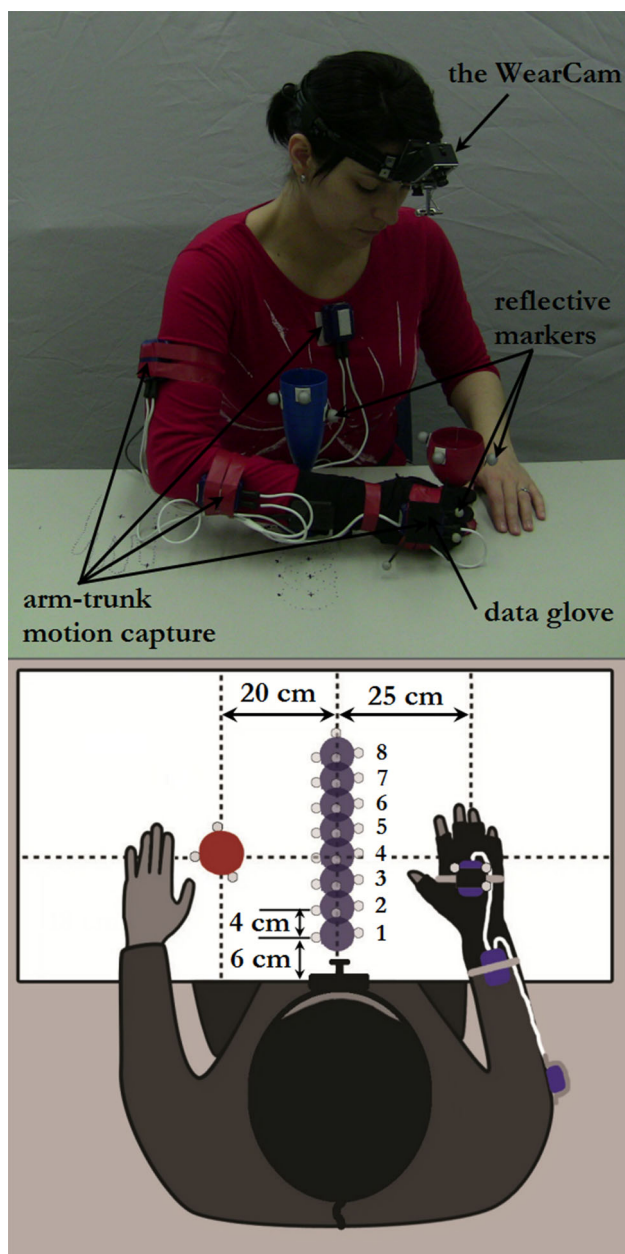
### 2.2 Experimental setup

Subjects sat in a height-adjustable chair facing a rectangular table with task-relevant objects placed on the surface of the table (Fig. 1). Subjects sat in front of a table such that the sagittal plane “cut” the width of the table at approximately the midline, and the distance from the frontal part of the trunk to the edge of the table was  $\sim 10$  cm. The initial positions of the right hand, the target object and the obstacle object were predetermined, and they were laid along a line parallel with the coronal plane of the body, 18 cm displaced from the edge of the table on the subject’s side. The distance measured in the table plane from the initial hand position (hand centroid) to the obstacle was 25 cm, and from the obstacle to the target, it was 20 cm (i.e., 45 cm from the starting hand position to the target). Starting positions were indicated by markers on the table. The two objects used for manipulation were IKEA glasses, color tinted to enable automatic color-based segmentation on video recordings. The wine glass (max. diameter 7.5 cm, height 13 cm) was the object to be grasped (target), and the champagne glass (max. diameter 5 cm, height 21 cm) was the object to be avoided (obstacle).

### 2.3 Task

Grasping during all trials was conducted with the right hand. The left hand remained on the table, to provide support for the trunk to reduce the movements of the trunk in the coronal plane. At the start of grasping, the subjects were instructed to look at the colored patch mounted on the data glove. A sound signal indicated the start of execution of grasping, instructing the subject that they were free to unlock gaze from the colored patch, mounted on the data glove, and start a trial. Once the grasping motion was completed, the subject was instructed to go to the starting position.

Each subject performed 8 trials of reaching for grasping the target (wine glass). In all the trials, the obstacle (champagne glass) was present. The location of the champagne glass was changed at each trial. Starting from 6 cm from the edge of the table on the subject’s side, we progressively displaced the champagne glass at each trial in increments of 4 cm along the midline of the desk (parallel to the sagittal



**Fig. 1** Experimental setup to record eye–arm–hand coordination from human demonstrations in grasping tasks where the obstacle (*dark blue disk*) is progressively displaced in each trial. Obstacle positions (superposed as transparent *dark blue disks*) are numbered from obs1–8, numbered with respect to the increasing distance from the subject. obs1 is the starting position of the obstacle, 6 cm from the edge of the table. We progressively displaced the champagne glass for each trial in increments of 4 cm along the midline of the desk. obs4 is the position of the obstacle which corresponds to the location of the object in the first experiment (18 cm from the edge), and obs8 is the farthest position of the obstacle (34 cm from the edge). In this trial, the human subject is grasping the target object (wine glass) avoiding the obstacle (champagne glass) (color figure online)

plane of the subject’s body in resting position). An alternative to this approach is to place the obstacle in a randomly indexed position for every trial. By incrementally displac-

ing the obstacle in each trial, we implicitly force subjects to change their previous obstacle avoidance strategy, whereas with random displacements, the hand path which assured successful obstacle avoidance in the previous trial (e.g., obstacle in position 4) could be reused for a new trial (e.g., obstacle in position 2), without much adaptation.

For all trials, subjects were instructed to perform manipulation in a natural manner, without any additional instructions that could affect their visuomotor behavior. The subjects had one trial of practice before recording to ensure that they had understood the instructions. Subjects were unaware of the purpose of the experiment. Figure 1 illustrates our setup for this experiment.

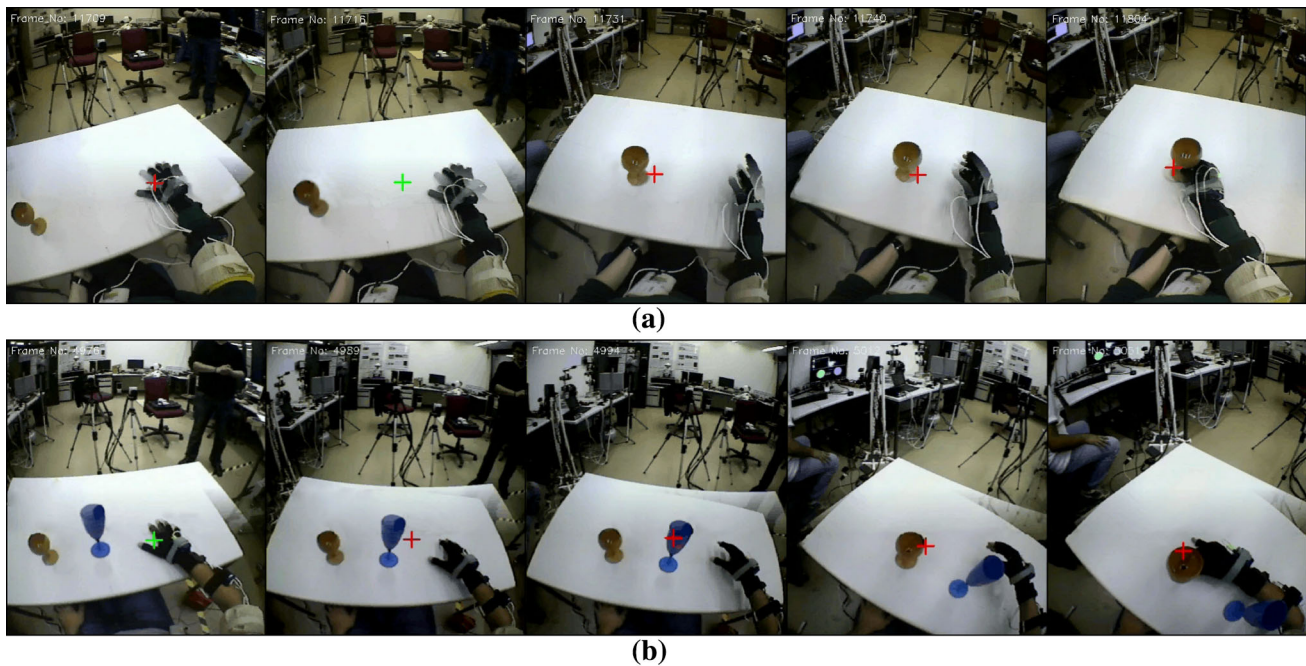
## 2.4 Apparatus

A head-mounted eye tracker designed in our laboratory, the WearCam system (Noris et al. 2010), was used for gaze tracking and for recording the scene as viewed from the subject’s standpoint. The system uses two CCD cameras to record a wide field of view ( $96^\circ \times 96^\circ$ ). It uses support vector regression to estimate the gaze direction from the appearance of the eyes. The system has an accuracy of  $1.59^\circ$ . The video and gaze positions from the WearCam were recorded in  $384 \times 576$  MJPEG format at 25 Hz. The WearCam video from our experiment can be seen in Fig. 2. The XSens<sup>TM</sup> inertial motion capture system was used for recording the trunk motion and arm motion. The sensors were mounted on the trunk, the upper arm, the forearm and the hand. The system provided information about three joints of trunk motion (roll, pitch and yaw), three joints that model the shoulder (flexion–extension, abduction–adduction and circumduction), two joints in the elbow (flexion–extension and pronation–supination) and two wrist angles (abduction–adduction and flexion–extension). The 5DT<sup>TM</sup> data glove, with flexure-sensors technology, was used for recording the finger joint angles motion. The data from the XSens<sup>TM</sup> IMU motion capture sensors and the 5DT<sup>TM</sup> data glove were recorded at 25 Hz.

The OptiTrack<sup>TM</sup> multi-camera system was used for tracking the 3D positions of the hand and the objects on the scene. The speed of data recording from the multi-camera system was 150 Hz, and the accuracy was  $\sim 2$  mm.

## 2.5 Calibration and data processing

The WearCam system was calibrated at the beginning and at the end of the task for each subject by using the procedure explained in Noris et al. (2010). The state of the WearCam was verified after each trial by checking its relative position with respect to the head and observing the video that was streamed. We checked the state of the multi-camera system by observing performances of real-time detection of the



**Fig. 2** Snapshots from the WearCam video from the start of the task (*left*) until the successful grasp completion (*right*), in a no-obstacle and **b** obstacle scenarios. The cross superposed on the video corresponds to

the estimated gaze position. The *color of the cross* indicates whether the gaze is the fixation state (*red*) or the saccade state (*green* (color figure online))

objects in the workspace, and we recalibrated it when the accuracy was not satisfactory. The data glove and the motion capture sensors were calibrated after each trial by requesting the subject to adopt an upright straight posture of the torso and to perform a sequence of opening and closing fingers. The state of the data glove and the motion capture sensors was verified by using an in-house GUI tool that shows body posture of the subject by using real-time readings from the sensors.

All recorded signals were filtered with a preprogrammed peak-removal technique that consisted in removing outliers from sensor misreadings and replacing them with linearly interpolated values between two closest valid readings. All signals were re-sampled at 25 Hz. The synchronization and parsing of signals were performed by using time stamps for recorded signals and verified by observing recorded videos on a frame-by-frame basis. The signals were smoothed with a moving average filter. Piecewise spline fitting was done, which did additional smoothing as well. Finally, we visually assessed comparative plots of both raw signals, and synchronized and smoothed signals in order to make sure that filtering and smoothing did not distort general signal profiles.

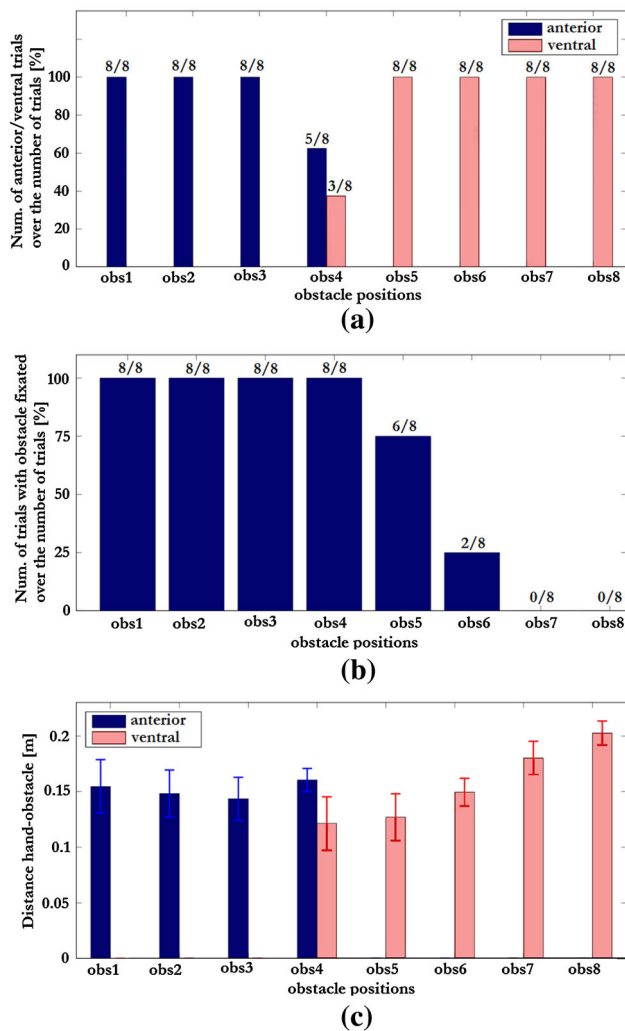
We detected gaze fixations as all instances where the gaze remained steady for at least 80 ms with gaze motion not exceeding  $1^\circ$  of the visual field (Inhoff and Radach 1998; Jacob and Karn 2003; Dalton et al. 2005). We say that a person is looking at either of the two objects (target or obstacle) if a gaze fixation is contained within the object blob, or it is within a 5-pixel radius around the object blob. This 5-pixel

radius accounts for imprecision in the blob segmentation, and in the estimation of the gaze position. It also accounts for the fact that the “functional fovea” forms a 3-degree circular region around the center of the gaze, which means that the visual system can obtain the high-quality visual information fixating very close to the edges of interesting objects (Rothkopf and Ballard 2009). We empirically obtained this specific value of a 5-pixel tolerance by computing the average closest distance between the estimated gaze point detected in the fixation state (but outside the segmented blob) and the boundary of the blob. This was done for a number of sub-parts of the reach-for-grasp task for which it is well-known that motoric actions impose strong demands for foveal visual information about the object’s state. One of the sub-parts of the task, when gaze fixations at the target object are expected with a high probability, is the moment just before the wine glass is grasped, as it is reported from previous studies that the gaze consistently fixates grasping parts before fingers touch the object (Brouwer et al. 2009).

## 2.6 Analysis of recordings from human trials

### 2.6.1 Visuomotor strategy and visuomotor coupling in obstacle avoidance

Figure 3a reveals the obstacle avoidance strategy that the subjects employed with respect to the position of the obstacle. It can be seen that the subjects preferred to avoid the obstacle



**Fig. 3** Results from the follow-up experiment with human subjects where the obstacle was progressively moved along the midline of the table: **a** influence of the position of the obstacle on strategy to avoid the obstacle from anterior/ventral side, **b** Influence of the obstacle position on gaze fixations at the obstacle during manipulation, and **c** Safety distances from the hand to the obstacle when avoiding it from anterior/ventral side

from the anterior side if the obstacle was positioned between the subject's body and the line that is defined from the starting position of the hand to the target object (obs1–4). If the obstacle was positioned in the anterior direction from the line (obs5–8), then the preferred obstacle avoidance strategy was to veer from the ventral side when reaching to grasp the target object. It can be seen that subjects are very consistent in their obstacle avoidance strategy, except for the obstacle position number 4 (obs4), for which 5 subjects avoided the obstacle from the anterior side and 3 subjects veered from the ventral side. Post hoc analysis of the recorded videos from the experiment revealed that 3 subjects who veered for obs4 from the ventral side kept the posture of torso more upwards than other subjects during manipulation, hence for them veering from

the ventral side was a choice that required less effort. Results presented here provide a basis for the computational model of our obstacle avoidance strategy regarding the choice of the preferred obstacle avoidance side, as discussed in Sect. 3.3.

An important part of the forward planning scheme is that an object in the workspace is tagged as an obstacle if it is estimated that the hand will collide with it. As the object identified as the obstacle is the intermediary target for the visuomotor system, it is expected that it will be visually fixated during reaching. Figure 3b shows the proportion of trials for each obstacle position in which the obstacle object was visually fixated. It can be seen that the champagne glass was always fixated when it was positioned on location 1 through 4 (obs1–4 in the figure). For position obs5, the obstacle was fixated in only 80 % of the trials. The amount of fixation rapidly drops to 20 % for position obs6, and to zero for positions obs7 and obs8. As expected, once the obstacle is sufficiently far, it is no longer of interest. These results are consistent with [Tresilian \(1998\)](#), who argued that objects treated as obstacles by the motor system are very likely to be visually fixated during manipulation. Thus, our results indicate that the most likely explanation of visual ignorance of the champagne glass when it is placed at obs6–8 is that the visuomotor planning scheme did not identify it as an obstacle.<sup>4</sup>

Based on the study by [Dean and Brüwer \(1994\)](#) and the results of our human experiment where the safety distance between the hand and obstacle was kept ([Lukic et al. 2012](#)), we hypothesized that the control system would keep the same safety margin of  $\sim 0.14 \pm 0.01$  m across all trials where the champagne glass was considered as an obstructing object (namely for position 1–6). In the other position, this safety margin would not be preserved as the obstacle would then be ignored.

In Fig. 3c, we plotted the minimum distance (the mean and the standard deviation) between the hand and the champagne glass for all positions of the champagne glass. It can be seen that the distance is quite consistent for obs1 to obs6 and starts increasing for obs7 and obs8. These results also indicate that an obstacle object positioned such that it does not obstruct the original prehensile motion is not identified as an obstacle and it is not treated as the intermediary target.

A two-way ANOVA (factors: subjects and a binary variable that represents whether the obstacle was fixated/not fixated in a trial) on the distance hand-obstacle reveals a significant effect of the *obstacle fixations* factor [ $F(1, 63) =$

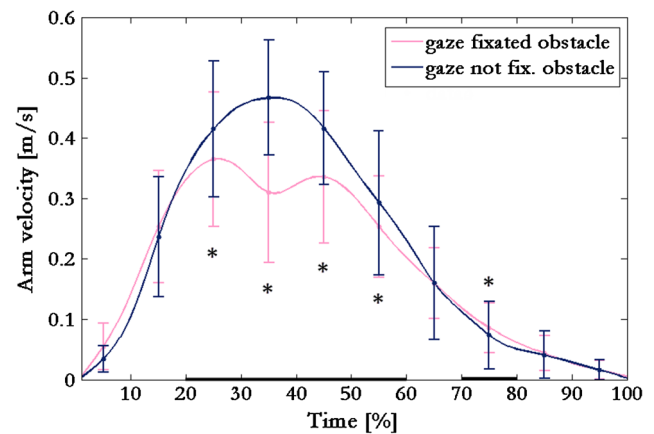
<sup>4</sup> At the end of all trials, we asked 2 subjects to try to reach for the target when the champagne glass (obstacle) was present, but without modification of the path (as in the no-obstacle setup). Unsurprisingly, the arm/hand collided with the champagne glass always when it was positioned at obs2, obs3, obs4, in 6 out of 8 trials the hand collided for obs1 and obs5. The hand never collided when the obstacle was in positions obs6, obs7 and obs8.



78.3,  $p < 0.001$ ], and no effect of the *subject* factor [ $F(7, 63) = 0.47$ ] and no factor interaction [ $F(7, 63) = 0.35$ ]. These results reflect the fact that the distance between the hand and the obstacle is significantly different when the subjects visually fixate the obstacle, compared to the case without gaze fixations at the obstacle object in a trial. We interpret these results as a confirmation of the influence of forward planning on visuomotor coordination. When forward planning estimates that the object obstructs intended movement, the motor system treats the obstacle as an intermediary target. The gaze fixates the obstacle, and the hand keeps a consistent safety distance from the object. If the object is placed in a position where it does not obstruct movements (obs6–8), it is not “tagged” as the obstacle. The visuomotor system ignores objects that are irrelevant to manipulation: They are not visually salient for the gaze (Land 1999; Hayhoe et al. 2003; Rothkopf et al. 2007; Rothkopf and Ballard 2009), and the hand is controlled without keeping some safety distance with respect to them.

We show that in the trials, where the location of the obstacle is varied, gaze fixations at the obstacle indicate that the arm keeps the safety distance from the obstacle. To further analyze the coupling between the gaze and the arm when performing obstacle avoidance, we investigated the influence of the gaze on the velocity profile of the arm. Alberts et al. (2002) and Hesse and Deubel (2010) showed that the velocity profile reaches usually a local minimum, when the arm passes the obstacle. In our experiment, the obstacle seems to influence the motion solely in trials when the gaze stops at the obstacle. We hence would expect that the motion of the arm would be slowed down at the obstacle only in these trials when the gaze fixates the obstacle. In the absence of the obstacle on the path toward the target, there should be no need to visually guide the arm to avoid it. Figure 4 compares the mean arm velocities across the trials in which the gaze fixated the obstacle versus the trials where the gaze did not fixate the obstacle. The observation of such a minimum velocity confirms the hypothesis that the obstacle acts as an intermediary target during movements (Alberts et al. 2002; Hesse and Deubel 2010). In contrast, and as hypothesized, the velocity profile in obstacle-free trials follows a regular bell-shaped profile.

We apply a two-way ANOVA on the velocity profiles recorded during trials with two factors: (a) an *obstacle fixations* factor representing the type of trial, coded as a binary variable, to distinguish between the conditions in which the obstacle was fixated versus not fixated; (b) a *time bin* index (the total time of each trial is divided into 10 equal time bins) to determine when, during a trial, an influence of the presence/absence of the obstacle could be observed. We observe a strong effect of the *obstacle fixations* factor [ $F(1, 6199) = 109.9$ ,  $p < 0.001$ ]. This confirms that the arm velocity profile is indeed significantly reduced when



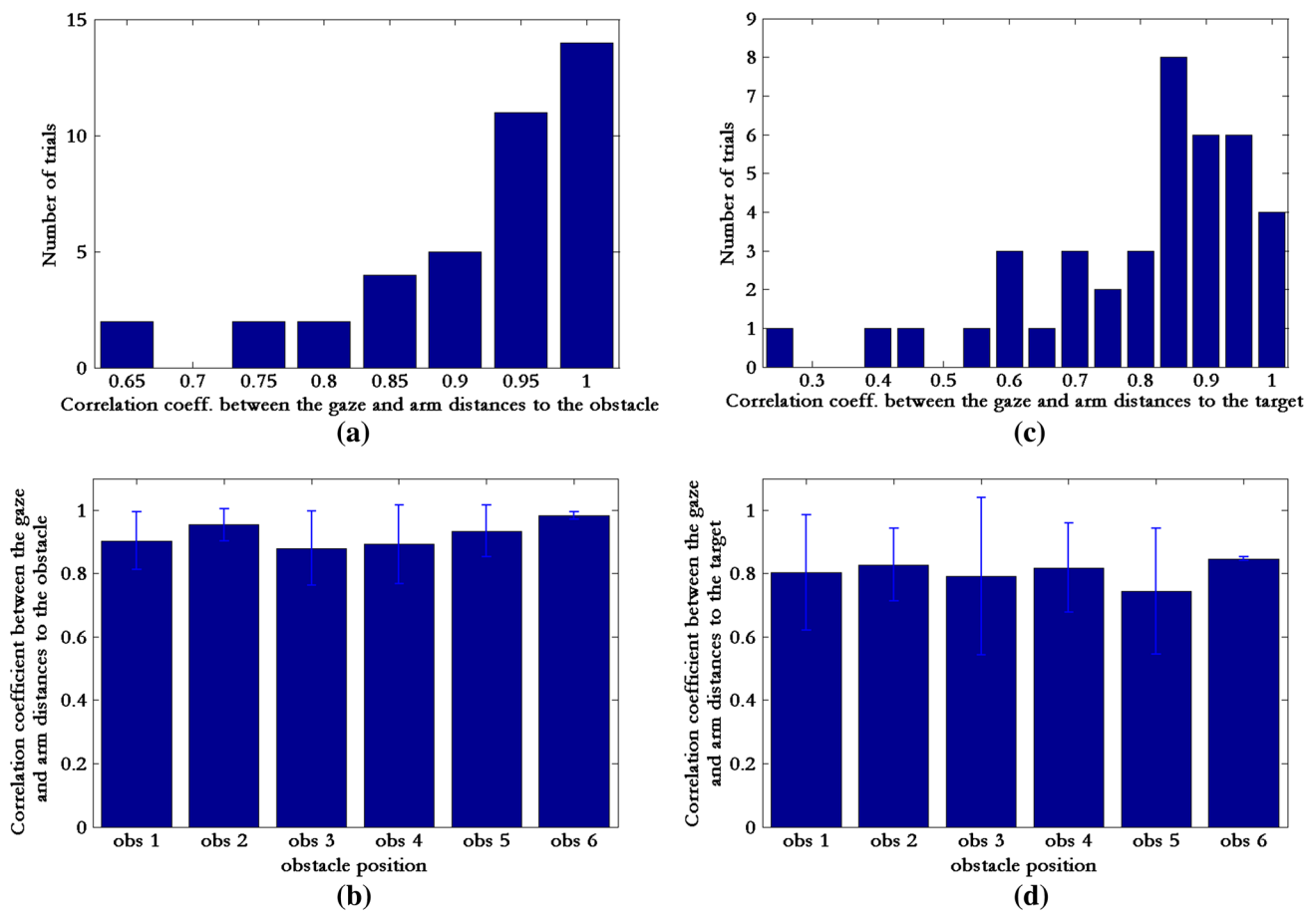
**Fig. 4** Arm velocity profiles, time normalized and averaged over all subjects for two conditions (gaze fixated the obstacle or not). Stars represent the time bins for which a post hoc  $t$  test shows significant difference between the fixation conditions ( $p < 0.05$ )

passing the obstacle. There is also a significant effect of the *time bin* factor [ $F(9, 6199) = 1849.44$ ,  $p < 0.001$ ], indicating that during the progress of the task arm velocity changed. As expected, the interaction between the factors is significant [ $F(9, 6199) = 41.44$ ,  $p < 0.001$ ] showing that the velocity profiles in trials where the gaze fixates the obstacle changes differently as the task progresses from the trials where the obstacle is not fixated. We run post hoc  $t$  tests between the fixated and not fixated trials to determine time bins for which the velocity arm profiles differ between the two conditions (Fig. 4).

The finding that the gaze fixations at the obstacle modulate the arm velocity profiles supports the hypothesis that the gaze–arm coupling exists when humans perform prehension with obstacle avoidance.

#### 2.6.2 Gaze–arm correlations

To see whether the gaze–arm mechanism follows a quasi-constant lag, we analyze trial-by-trial correlations between the gaze and arm positions (computed as the Euclidean distance) with respect to the obstacle (in the first segment of the movement) and correlations between the gaze and arm distances with respect to the target (in the second segment of the movement) as the task progresses. We plot the histogram of the Pearson’s correlation coefficient between the gaze and the arm distances computed on a trial-by-trial basis when approaching to the obstacle (Fig. 5a, b) and the target (Fig. 5c, d). We see the prevalence of very high visuomotor correlations for both objects. The distribution of trial-by-trial correlation coefficient between the gaze and arm distances to the obstacle has a sample mean of 0.917, and the 25, 50 % (median) and 75 % percentile correspond to 0.876, 0.956 and 0.986, respectively. Similarly, the correlation coefficient



**Fig. 5** Correlation coefficient between the gaze and arm distances with respect to the obstacle and the target computed on a trial-by-trial basis when avoiding the obstacle. The motion is segmented into two parts: from the starting position to the obstacle and from the obstacle to the target and we compute the correlations for the corresponding parts of the

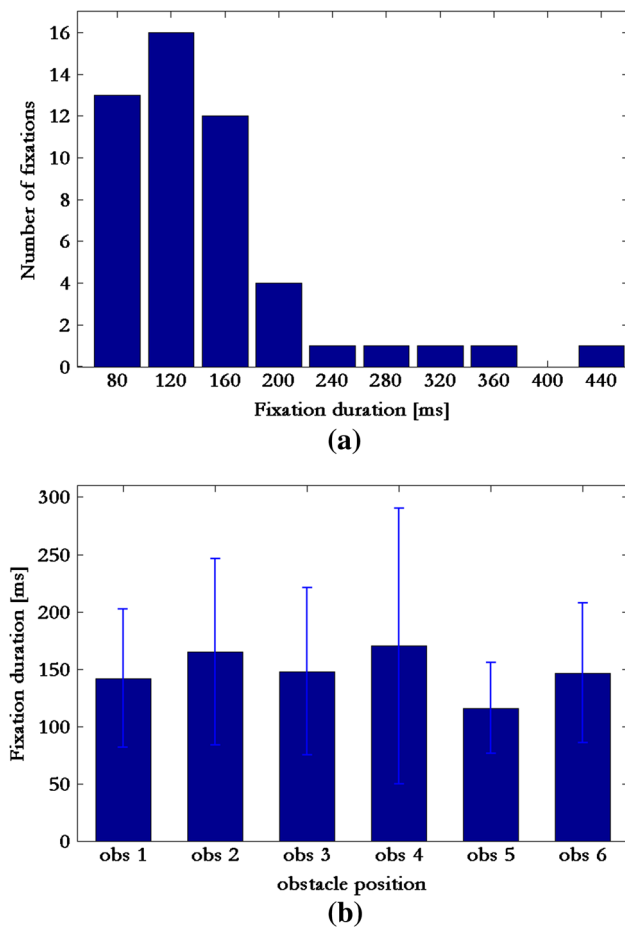
movements: **a** histogram of the gaze–arm correlation coefficient when reaching the obstacle and **b** corresponding values for different fixated obstacle positions, **c** histogram of the gaze–arm correlation coefficient when reaching the target and corresponding values for different fixated obstacle positions (**d**)

between the gaze and arm distances to the target has the sample mean 0.799, and the 25, 50 % (median) and 75 % percentile correspond to 0.721, 0.847 and 0.921, respectively. A two-way ANOVA for the correlations to the obstacle (factors: subjects and obstacle position) does not reveal a statistical significance of the *subject* factor ( $p = 0.186$ ) and no effect of the *obstacle position* factor ( $p = 0.77$ ). A two-way ANOVA for the correlations to the target (factors: subjects and obstacle position) shows no statistical significance *subject* ( $p = 0.164$ ) and no effect of the *obstacle position* ( $p = 0.934$ ) as well.

The correlations between the gaze and arm trajectories when reaching for the obstacle are quasi-constant across trials and subjects, and they are almost the same as those observed for the target. These observations suggest that the eyes and the arm might be driven to both the obstacle and the target by the same mechanism of spatial coordination.

### 2.6.3 Fixation durations at the obstacle

We now present the results of our analysis of gaze fixation durations at the obstacle. It is well established that the gaze fixation durations, together with the position of the gaze, provide a measure of cognitive processing when performing an ongoing task, being positively correlated with cognitive load required for processing visual information (Rayner 1998; Deubel et al. 2000; Jacob and Karn 2003; Hayhoe and Ballard 2005; Tatler et al. 2011). Gaze fixations in visually guided manipulation allow very specific task-dependent acquisition of visual information (Triesch et al. 2003). This selectivity in information processing is reflected in the duration of fixations (i.e., a variability in fixation duration corresponds to a variability in visual features being selectively acquired from the early visual structures and further processed in the higher cortical structures). Figure 6a shows the histogram of the



**Fig. 6** Distribution of gaze fixation durations at the obstacle: **a** histogram of fixation durations pooled from all subjects across all fixated obstacle positions, **b** the mean and the standard deviations of times for different fixated obstacle positions. In this plot, we show only fixations times and the standard deviations for positions at which the obstacle is fixated (obs1–6), and positions obs7, obs8 are omitted from the figure because subjects never fixated the obstacle when it was placed at these positions

fixation durations at the obstacle where the data are pooled from all subjects. The distribution is positively skewed with the sample mean fixation duration at 146.4 ms, where the 25, 50 % (median) and 75 % percentile correspond to 80, 120 and 160 ms, respectively. The predominance of short fixations observed in our experiment is a common feature of a gaze fixation pattern in natural manipulation tasks (Land 1999; Hayhoe et al. 2003; Hayhoe and Ballard 2005), where the average durations of fixations are shorter compared with durations observed in picture viewing and reading (Rayner 1998; Henderson and Hollingworth 1999). In spite of the predominance of brief durations of fixations in prehension movements, it has been shown that they do support movement control. Several studies have shown that visual information necessary for movement control can be computed within a single fixation (Ballard et al. 1995; Land et al. 1999). This

indicates quite efficient visual processing of some easy-to-compute visual features required for online arm movement control. A two-way ANOVA (factors: subjects and an index variable that represents a position of the obstacle) shows no significant effect of *subject* factor ( $p = 0.321$ ) and no effect of the *obstacle position* factor ( $p = 0.564$ , see Fig. 6b) indicating that fixations times are consistent both across subjects and obstacle positions. These results are in agreement with the prior results of Johansson et al. (2001) who observed the predominance of brief fixations at the obstacle. An interesting result comes from one of their obstacle avoidance experiments. When active gaze movements were inhibited during obstacle avoidance, they observed a great variability in the minimum distance kept between the obstacle and the hand. We can speculate that the existence of these brief and quite consistent fixation times reflect the consistency in processing simple visual features of the obstacle in order to guide the arm and hand, because the existence of brief fixation periods does not allow to compute some complex features such as in reading (Rayner 1998). Considering the predominance of brief fixation times and an increased variability of estimating the position of the obstacle, one of these features computed is most likely the spatial position of the obstacle. The spatial location of the obstacle can be rapidly computed from retinal (foveal and parafoveal visual information) and extraretinal information (the relative position of the eyes and the head) available at the moment of fixation by the specialized neural circuitry of the dorsal visual stream (Goodale and Haffenden 1998; Goodale 2011), and it is a necessary feature in order to safely guide the arm around the obstacle.

In summary, this analysis of the duration of the gaze fixations provides support to the view that the CNS computes simple features during fixations at the obstacle in order to aid obstacle avoidance. The spatial location of the obstacle is likely one of the main features computed during these gaze fixations on the obstacle.

#### 2.6.4 Gaze and arm exit times from the obstacle

We provide a quantitative assessment of the relation between the gaze exit time and the arm exit time from the obstacle.<sup>5</sup> If some coordination exists between the gaze and the arm when performing obstacle avoidance, these two measures should be correlated. Moreover, the magnitude of the lag between them (i.e., the difference between the exit times of the gaze and arm from the zone of the obstacle) should be kept relatively tight compared with the overall time necessary to

<sup>5</sup> The gaze exit time from the obstacle is defined as the time from the beginning of a trial until the onset of a saccade away from the fixated obstacle. The arm exit time is defined as the time from the beginning of a trial until the moment when the arm reaches the closest distance to the obstacle and starts moving toward the target.

complete the movement. When plotting the onset time of the gaze versus the arm onset time from the obstacle pooled from all subjects (except for Subject 1<sup>6</sup>), we can see from Fig. 7a that these two variables are linearly correlated (Pearson's correlation coefficient  $r = 0.897$ ,  $p < 0.001$ ). The slope of the fit indicates that, on average, the gaze exits the obstacle zone slightly earlier than the hand. Figure 7b shows the histogram of the difference between the gaze exit times and arm exit times, where positive values indicate that the gaze exits the obstacle first. The distribution has the sample mean at 220.78 ms, where the 25, 50 % (median) and 75 % percentile correspond to 120, 200 and 280 ms, respectively. A two-way ANOVA (factors: subjects and an index variable that represents a position of the obstacle) shows no significant effect of *subject* factor ( $p = 0.18$ ) and no effect of the *obstacle position* factor ( $p = 0.549$ ), indicating that the difference between gaze and arm exit times was consistent both across subjects and obstacle positions (Fig. 7c). The predominance of positive differences gives evidence that the gaze leaves the obstacle before the hand leaves it. However, the median time of this lag corresponds to only 8.3 % of the median time (2.4 s) needed to complete the whole reaching movement with obstacle avoidance. This means that this period of apparent asynchrony after the gaze switched toward the target while the arm is in the obstacle zone takes only a small fraction of the overall movements. For the remaining 91.7 % of the task, gaze and arm movements are synchronously driven to the same goal (to the obstacle during the first segment of movement, and toward the target after the obstacle is passed). Land et al. (1999) observed in their team-making experiment that the gaze and arm movements are highly coupled during execution of each subtask, but when it comes to a transition toward a new target, the gaze switches approximately 0.5 s before the movement of the arm to the previous object is completed. Johansson et al. (2001) found that the difference between the gaze exit times and arm exit times was quite tight when executing sequential tasks, but the gaze starts moving toward the new target slightly before the hand does (~100–200 ms), as well. The results were similar

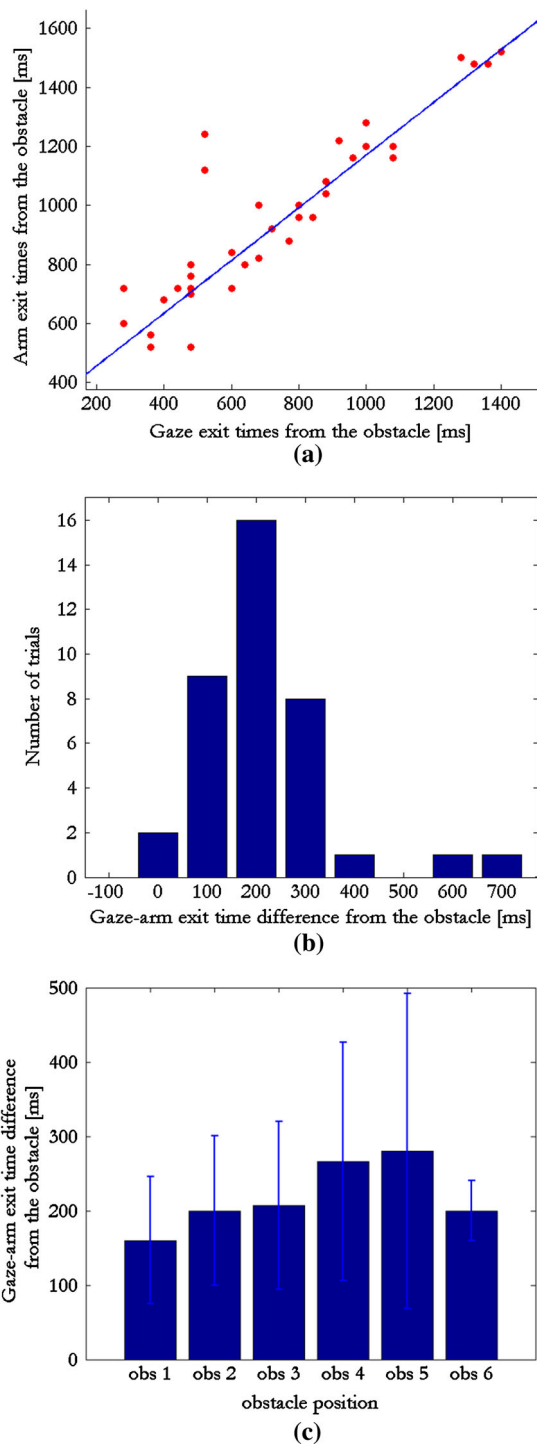
for a number of different movement sub-targets, including the obstacle.<sup>7</sup>

From our results and from the two aforementioned studies, it is evident that the gaze and the arm exit times when completing one movement segment and switching to a new target are tight compared with the average duration of movements. Nevertheless, it remains to be discussed why this lag is not exactly zero meaning that the gaze and the arm switch to the next target at exactly the same time. We here provide two alternative explanations.

First, this lag may be due solely to the well-known delays in processing the visuomotor control loop. Such delays are of the order of 100–250 ms (Wolpert et al. 1998, 2001), which amounts to the time delays in our experiments. Although the dorsal visual stream is capable of performing fast visuomotor transformations, it is possible that switching toward the new target is easier for the gaze than for the arm, due to both the grater physiological complexity of the arm control system and increased delays resulting from longer neural pathways. However, one could state an alternative explanation that relates to the fundamental control strategy in the CNS. Because the arm avoids the obstacle at some safety distance, and the experimental task is designed such that obstacle position is kept constant during the trials, the “buffered” position of the obstacle from the last fixation at the obstacle is a very good reference point for the arm. Land and Furneaux (1997) have shown that information buffering of spatial coordinates acts as an adjunctory mechanism when transitions between visuomotor sequential tasks occur. The arm is at the moment when the gaze leaves the obstacle displaced at some distance from to the obstacle and hence neither much adjustment is needed nor very precise visual information is needed to avoid the obstacle. This could be an efficient strategy in terms of the attentional resources considering that there is neither much surprise in the task nor the extreme precision is required. This suggests that the CNS employs “loose” transition between the subtasks, saving valuable, limited attentional resources, whenever prior information about the task suggests that not much change in the workspace is

<sup>6</sup> The coordination of the gaze and arm exit times from the obstacle for Subject 1 substantially differed from the rest of the subjects. She has shown significantly different amount of the gaze–arm lag when exiting the zone of the obstacle (mean 448 ms, SD 210.5 ms) compared to the rest of the subjects (mean 220.78 ms, SD 135.75 ms) and this difference achieved statistical significance [one-way ANOVA:  $F(1, 39) = 10.93$ ,  $p = 0.002$ ]. A careful analysis of the video from the eye tracker revealed her visuomotor strategy. Interestingly, her eye and arm movements were normal and the gaze guided the arm in all trials. However, she mostly used the coordination strategy where the gaze first visits the obstacle and the moment when gaze switches toward the target she started to move the arm, i.e., start of her arm movement was significantly postponed. In all the other measures she did not significantly differ from the rest of the subjects.

<sup>7</sup> It is important to note that Johansson et al. (2001) focused most of their analysis on gaze and arm timing with respect to entering or exiting the so-called landmark zones. They defined the landmark zone as an area with the radius 3° of visual angle (2 cm) in the work plane in all directions from the corresponding objects in the workspace, including the obstacle. They found that the gaze and arm have almost identical exit times from the obstacle landmark zone. Considering that an approximate overall vertical arm displacement in their experiment was 12 cm, these landmark zones established a coarse representation of the workspace. However, from the plots where precise spatio-temporal measures were presented (Fig. 6A in their paper), it can be seen that the difference between the median gaze and arm exit times at the exact location of the obstacle differs approximately 200 ms in favor of gaze exiting first the obstacle. Similar measures of the gaze–arm exit lag hold for the other intermediary targets (e.g., support surface, target switch and bar tool).



**Fig. 7** Gaze exit times versus arm exit times from the obstacle: **a** scatter plot of gaze exit times versus arm exit times from the obstacle pooled from Subjects 2–8 across all fixated obstacle positions, **b** histogram of gaze–arm exit time differences from Subjects 2–8 across all fixated obstacle positions, where positive values mean that the gaze exits the obstacle zone before the arm, **c** the mean and the standard deviations of gaze–arm exit time differences for different fixated obstacle positions. In this plot, we show only fixations times and the standard deviations for positions at which the obstacle is fixated (obs1–6), positions obs7–8 are omitted from the figure because subjects never fixated the obstacle when it was placed at these positions

expected and not much accuracy is needed. In the task where sequential movements had very high precision constraints by means of the requirements of precisely touching a target, the gaze exit times were almost always tightly synchronized with the arm exit times (Bowman et al. 2009). The experiment of Bowman et al. (2009) shows that the “tight” switching strategy holds as well.

This analysis shows that the gaze exit times and arm exit times from the obstacle are highly correlated, suggesting strong visuomotor synchronization with respect to the obstacle. The time difference between the gaze and the arm times when switching from the obstacle is nonzero positive, but it remains small compared with the overall task duration.

### 2.6.5 Summary

In summary, the mechanism of the eyes leading the arm is observed in all trials. This study corroborates other findings in the literature on a strong coupling between arm and eye motion, where the eyes lead the arm in a systematic and coordinated pattern. Additionally, it supports the hypothesis that the obstacle may act as an intermediary target. We should emphasize that this study is particularly instrumental in providing us with quantitative data onto which to ground the parameters of the model, as we describe next.

## 3 Computational approach and system architecture

In the first part of this section, we introduce the principle of robot control by using time-invariant DSs and the probabilistic approach for estimating the parameters of the system. Furthermore, we extend this formulation for modeling and control of coupled dynamics. Finally, we show how the basic model of eye–arm–hand coordination in the obstacle-free grasping can be extended to handle the obstacle in the workspace.

### 3.1 A single DS and GMM/GMR

The motion of our system is represented through the state variable  $\xi \in \mathbb{R}^d$ , symbolizing retinal coordinates representing the gaze state, Cartesian coordinates for the arm state and finger joint angles for the hand state.  $N$  recorded demonstrations of the task yield the data set  $\{\xi_t^n, \dot{\xi}_t^n\}, \forall t \in [0, T_n]; n \in [1, N]$ , of the robot’s states and state derivatives at particular time steps  $t$ , where  $T_n$  is the number of samples in the  $n$ th demonstration. We posit that the recorded data samples are instances of motion governed by a first-order autonomous differential equation:

$$\dot{\xi} = f(\xi) + \epsilon \tag{1}$$

where  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a continuous and continuously differentiable function, with a single equilibrium point  $\xi^* = f(\xi^*) = 0$ .  $\epsilon$  is a zero-mean Gaussian noise. The noise term encapsulates both sensor inaccuracies and errors inherited from human demonstrations. Time invariance provides inherent robustness to temporal perturbations. In order to achieve robustness to displacement in the position of the target, the robot’s state variable  $\xi$  is represented in the target’s reference frame.

We use the Gaussian mixture model (GMM) to encode dynamics in a probabilistic framework. The GMM defines a joint probability distribution function  $\mathcal{P}(\xi_t^n, \dot{\xi}_t^n)$  over the set of data from demonstrated trajectories as a mixture of  $K$  Gaussian distributions (with  $\pi^k, \mu^k$  and  $\Sigma^k$  being the prior probability, the mean value and the covariance matrix of the  $k$ th Gaussian, respectively):

$$\mathcal{P}(\xi_t^n, \dot{\xi}_t^n) = \sum_{k=1}^K \pi^k \mathcal{N}(\xi_t^n, \dot{\xi}_t^n; \mu^k, \Sigma^k), \tag{2}$$

where each Gaussian probability distribution is defined as:

$$\begin{aligned} \mathcal{N}(\xi_t^n, \dot{\xi}_t^n; \mu^k, \Sigma^k) &= \frac{1}{\sqrt{(2\pi)^{2d} |\Sigma^k|}} e^{-\frac{1}{2} (([\xi_t^n, \dot{\xi}_t^n] - \mu^k)^T (\Sigma^k)^{-1} ([\xi_t^n, \dot{\xi}_t^n] - \mu^k))} \end{aligned} \tag{3}$$

where the mean and the covariance matrix are defined as:

$$\mu^k = \begin{pmatrix} \mu_{\xi}^k \\ \mu_{\dot{\xi}}^k \end{pmatrix} \text{ and } \Sigma^k = \begin{pmatrix} \Sigma_{\xi\xi}^k & \Sigma_{\xi\dot{\xi}}^k \\ \Sigma_{\dot{\xi}\xi}^k & \Sigma_{\dot{\xi}\dot{\xi}}^k \end{pmatrix}. \tag{4}$$

We use the stable estimator of dynamical systems (SEDS) (Khansari-Zadeh and Billard 2011) to compute the GMM parameters. The SEDS ensures global stability of the noise-free estimate of the underlying dynamics, denoted as  $\hat{f}$ .

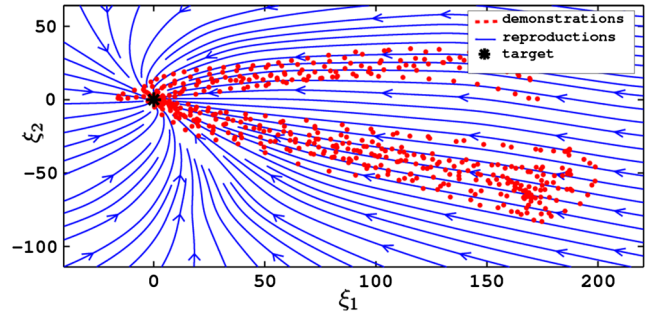
Taking the posterior mean estimate of  $\mathcal{P}(\xi_t^n | \xi_t^n)$  yields an estimate of  $\dot{\xi} = \hat{f}(\xi)$ , a function that approximates the model dynamics through a mixture of  $K$  Gaussian functions:

$$\dot{\xi} = \sum_{k=1}^K h^k(\xi) (A^k \xi + b^k), \tag{5}$$

where  $h^k(\xi), A^k$  and  $b^k$  are defined as:

$$\begin{cases} h^k(\xi) = \frac{\pi^k \mathcal{N}(\xi; \mu^k, \Sigma^k)}{\sum_{i=1}^K \pi^i \mathcal{N}(\xi; \mu^i, \Sigma^i)} \\ A^k = \Sigma_{\xi\xi}^k (\Sigma_{\dot{\xi}\xi}^k)^{-1} \\ b^k = \mu_{\dot{\xi}}^k - A^k \mu_{\xi}^k. \end{cases} \tag{6}$$

A toy example with a 2-dimensional DS, which illustrates the principles of encoding the demonstrated motion and robot control by using a time-invariant DS, is presented in Fig. 8.



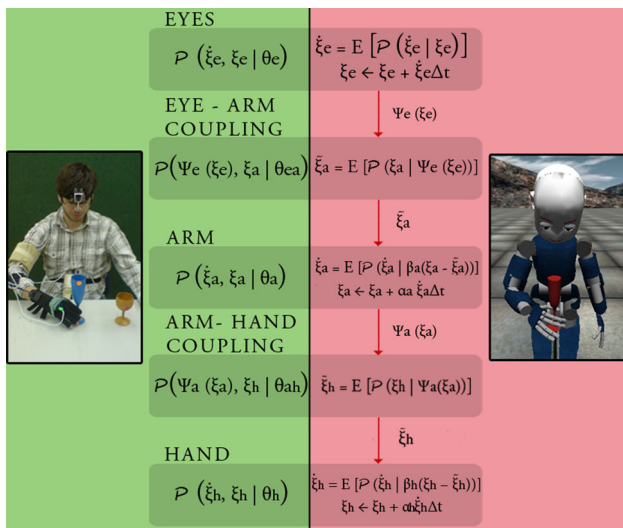
**Fig. 8** Learning and reproducing a motion with a single time-invariant DS. Given a set of demonstrations (red points), we build an estimate of an underlying dynamics. The asymptotic stability of the DS guarantees that the target (black star) will be reached. The DS, for a given robot state, computes a velocity vector that moves the robot state toward the target; hence, it can be illustrated with streamlines (blue lines) in the state space that steer the robot state toward the target (color figure online)

### 3.2 Coupled dynamical systems

Our recent work (Shukla and Billard 2011) shows the benefits of explicitly learning a coupling between the arm DS and the finger DS over modeling motions of the physical systems with a single extended DS. The problem associated with learning one high-dimensional dynamical model that guides the motion of two physical systems is that an explicit following of correlations shown in demonstrations between the two coupled dynamics is not guaranteed. This could be a problem if the robot is perturbed far from the region of demonstrated motion, as the behavior of the dynamical systems may not be correctly synchronized. The loss of coordination between the reach and grasp components might lead to failure of the overall prehensile task even when the individual dynamical systems converge to their attractors. An approach adopted in Shukla and Billard (2011) is to learn separately two dynamics and then learn a coupling between them. This approach ensures that the two DS will converge to their attractors, following a learned pattern of coordination between them. The approach, where the arm and hand DS are learned separately and then coupled explicitly, ensures that the behavior of the two systems is correctly synchronized, even when the motion is abruptly perturbed far from the motion recorded in human demonstrations. For more details about general properties of CDS, see Shukla and Billard (2011).

#### 3.2.1 Extended CDS architecture and learning

We extend the original CDS architecture with in total five building “blocks”: three dynamical systems and two coupling blocks between them. They are organized in the following order: eye dynamics → eye–arm coupling → arm dynamics → arm–hand coupling → hand dynamics, where



**Fig. 9** CDS-based robotic eye–arm–hand coordination. *Left (green part)* of the figure shows how the CDS model is learned. Reproduction of motion on the robot is shown on the *right side* of the figure (*red part*). CDS consists of five building “blocks”: three dynamical systems (the eyes, the arm and the hand) and two coupling models: eye–arm coupling and arm–hand coupling (color figure online)

the arrow direction indicates the direction of control signals. The gaze DS is the master to the arm DS, and the arm DS is the master to the hand DS. There is a coupling block between each master and its slave. The major assumption is that the modulation signals between them flow only in the direction from the master to the corresponding slave, i.e., the dynamics of the slave is modulated with control signals coming from its master, not vice versa. The master system evolves independently of its slave. Figure 9 illustrates the architecture of CDS, and the principles of learning and of the reproduction of coordinated motion.

The state of the eyes is denoted with  $\xi_e \in \mathbb{R}^2$ , the state of the arm is  $\xi_a \in \mathbb{R}^3$ , and the state of the hand is  $\xi_h \in \mathbb{R}^9$ . The eye state  $\xi_e$  is represented as the distance between the position of the gaze and the position of a visual target in retinal coordinates (i.e., retinal error). The arm state  $\xi_a$  is represented as the distance in Cartesian coordinates between the palm center and the final palm position with respect to the target object. The hand state  $\xi_h$  is expressed as the difference between the current hand configuration and the goal hand configuration, i.e., hand configuration adopted when the target object is in the grasp. In other words, the attractors of the eye, arm and hand DS are placed at the target projection in the retinal plane, its Cartesian position in the workspace and at the corresponding hand configuration when the target is grasped, which is formally expressed as:  $\xi_e^* = 0$ ,  $\xi_a^* = 0$  and  $\xi_h^* = 0$ , respectively.

Our CDS model of eye–arm–hand coordination is built in the following manner. We first learn separately joint probability distributions that encode the eye dynamics  $\mathcal{P}(\xi_e, \xi_e | \theta_e)$ ,

the arm dynamics  $\mathcal{P}(\xi_a, \xi_a | \theta_a)$  and the hand dynamics  $\mathcal{P}(\xi_h, \xi_h | \theta_h)$ . Then we learn the joint distribution for eye–arm coupling  $\mathcal{P}(\Psi_e(\xi_e), \xi_a | \theta_{ea})$  and arm–hand coupling  $\mathcal{P}(\Psi_a(\xi_a), \xi_h | \theta_{ah})$ , where  $\theta_e, \theta_a, \theta_h, \theta_{ea}$  and  $\theta_{ah}$  denote the GMM parameters, and  $\Psi_e(\xi_e)$  and  $\Psi_h(\xi_h)$  denote the coupling functions. GMMs that encode the dynamics of the eyes, the arm dynamics and the hand dynamics are learned using the SEDS algorithm, for more details see [Khansari-Zadeh and Billard \(2011\)](#). GMMs that model eye–arm and arm–hand coupling are learned with the expectation-maximization (EM) algorithm ([Bishop 2007](#)).

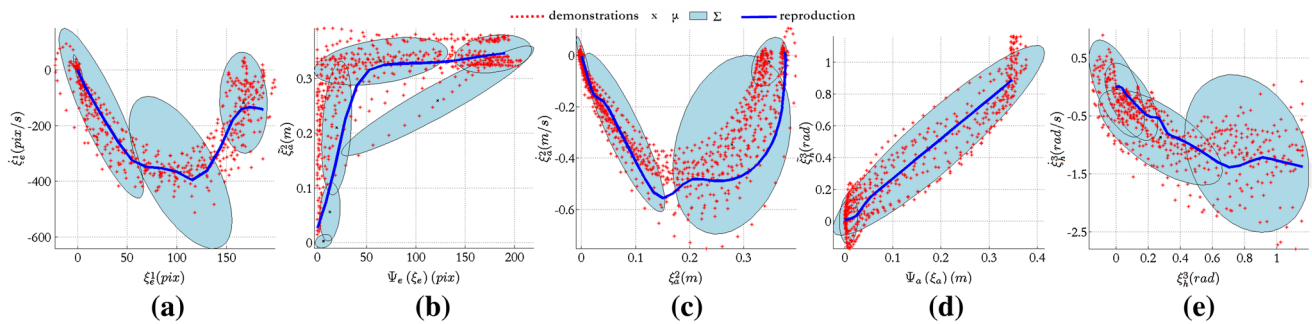
Two open parameters,  $\alpha$  and  $\beta$ , allow for an additional fine-tuning of the characteristics of the slave response ( $a$  and  $h$  subscripts denote whether they modulate arm or hand motion, respectively). The speed is modulated with the scalar  $\alpha$ , and the amplitude of motion is tuned by changing the value of the scalar  $\beta$ . Some robots can move faster than humans, hence by using larger values for  $\alpha_a$  and  $\alpha_h$ , one can exploit the robot’s fast reaction times. One can tailor the amplitudes of reactions to perturbations, suitable for a robot platform and a given task, by modulating the values of  $\beta_a$  and  $\beta_h$ .

Figure 10 illustrates the CDS model learned from demonstrations.

### 3.2.2 CDS reproduction

The DS that drives the eyes evolves independently in time and leads the whole system. The eye state velocity  $\dot{\xi}_e$  is generated by conditioning the eye dynamics model on the current eye state. The learned GMMs are conditioned by computing the Gaussian mixture regression (GMR) function (Eq. 5), for more about GMR see [Sung \(2004\)](#). The eye state variable is incremented by adding the computed velocity multiplied by the time step  $\Delta t$  to its current value  $\xi_e$ . The desired arm state value  $\tilde{\xi}_a$  is inferred from the eye–arm coupling model by conditioning on the eye–arm coupling function  $\Psi_e(\xi_e)$ . The arm velocity  $\dot{\xi}_a$  is computed by conditioning the arm dynamics model on the difference between the current and desired value  $\xi_a - \tilde{\xi}_a$ . The arm state variable is incremented by adding the computed velocity multiplied by  $\Delta t$  to its current value  $\xi_a$ . The desired hand state value  $\tilde{\xi}_h$  is obtained by conditioning the arm–hand coupling model on the arm–hand coupling  $\Psi_a(\xi_a)$ . The hand velocity  $\dot{\xi}_h$  is inferred by conditioning the hand dynamics model on  $\xi_h - \tilde{\xi}_h$ . Finally, the hand state variable is incremented by adding the computed velocity multiplied by  $\Delta t$  to its current value  $\xi_h$ . The eyes, arm and hand reach commanded states, and the loop is repeated until the target object is grasped.

Algorithm 1 shows how the robotic eye–arm–hand coordination is performed with CDS.



**Fig. 10** Learned CDS eye–arm–hand coordination model: **a** eye dynamics, **b** eye–arm coupling, **c** arm dynamics, **d** arm–hand coupling and **e** hand dynamics. For simplicity of graphical representation, we plotted the CDS model for one gaze position, one arm position and one hand position. The eye state is presented with horizontal gaze coordinate, denoted as  $\xi_e^1$ . The arm state is presented with Cartesian coordinate

### Algorithm 1 CDS eye–arm–hand coordination

do

  General:

- query frames from cameras
- read the current hand position from forward kinematics
- read the hand joints from encoders
- recognize and segment the target object
- estimate the position of the target in both retinal and Cartesian coordinates
- compute  $\xi_e$ ,  $\xi_a$  and  $\xi_h$

  Gaze:

**if** gaze is not at target **then**

$$\dot{\xi}_e \leftarrow \mathbb{E}[\mathcal{P}(\dot{\xi}_e | \xi_e)]$$

$$\xi_e \leftarrow \xi_e + \dot{\xi}_e \Delta t$$

- look at new gaze point

**end if**

  Eye – arm coupling:

$$\dot{\xi}_a \leftarrow \mathbb{E}[\mathcal{P}(\dot{\xi}_a | \Psi_e(\xi_e))]$$

  Arm:

**if** the arm is not at target **then**

$$\Delta \xi_a \leftarrow \xi_a - \dot{\xi}_a$$

$$\dot{\xi}_a \leftarrow \mathbb{E}[\mathcal{P}(\dot{\xi}_a | \beta_a \Delta \xi_a)]$$

$$\xi_a \leftarrow \xi_a + \alpha_a \dot{\xi}_a \Delta t$$

- solve inverse kinematics
- move the arm and the torso to new joint conf.

**end if**

  Arm – hand coupling:

$$\dot{\xi}_h \leftarrow \mathbb{E}[\mathcal{P}(\dot{\xi}_h | \Psi_a(\xi_a))]$$

  Hand:

**if** the hand is not at target **then**

$$\Delta \xi_h \leftarrow \xi_h - \dot{\xi}_h$$

$$\dot{\xi}_h \leftarrow \mathbb{E}[\mathcal{P}(\dot{\xi}_h | \beta_h \Delta \xi_h)]$$

$$\xi_h \leftarrow \xi_h + \alpha_h \dot{\xi}_h \Delta t$$

- move the hand to new joint conf.

**end if**

**until** object grasped

that corresponds to the direction of the major hand displacement in the task, denoted as  $\xi_a^2$ . The hand state is represented with thumb proximal joint, denoted as  $\xi_h^3$ . Superposed to the datapoints, we see the regression signal (*plain line*) and the different Gaussian distributions (*elliptic envelopes*) of the corresponding Gaussian Mixture Models

acts as the intermediary target for the visuomotor system in reaching and grasping tasks, see Sect. 2.

In order to define which objects in the workspace are obstacles for the realization of the intended reach and grasp tasks, we use a planning scheme to estimate the consequences of future actions. More specifically, the motion of the arm toward the target is estimated by integrating the dynamics of the extended CDS until each DS reaches its attractor. We integrated only the eye–arm part of the whole CDS, ignoring the hand’s DS, as our collision checking scheme is relatively simple. The arm end-effector is modeled as a point that moves along the estimated trajectory. Obstacle objects in the workspace are modeled as cylinders. The dimensions of a modeling cylinder should enclose the actual dimensions of the object, but should also account and compensate for the fact that the hand was modeled as a point. This is achieved by expanding the modeling cylinder for some predetermined fixed distance (we used 5 cm for both radius and height) from the dimensions where it fits exactly around the object. By taking this approach, we are able to reliably detect collisions with the fingers in our forward planning scheme, even though the hand is modeled as a point. The argument for using this simplistic collision checking scheme is our attempt to minimize additional computational load in the control loop.

An object is tagged as an obstacle when the trajectory of the end-effector intersects with a cylinder modeling the object (certain collision), or when the cylinder lies within the area where it is very likely that it will collide with the forearm (very likely collision). For the motions, we consider here and by observing the iCub’s body, we define this area as the slice of the workspace enclosed by the estimated trajectory of the end-effector and the coronal plane of the body.

As suggested earlier on, we consider the eye–arm–hand coordination as a composition of two segments: a motion from the starting position toward the obstacle and from the obstacle toward the target object. Individual segments of

### 3.3 Eye–arm–hand coordination for obstacle avoidance

The extension of the CDS eye–arm–hand controller for obstacle avoidance is grounded on our hypothesis that the obstacle



coordinated motion (from the starting point to the obstacle, and from the obstacle to the target) are performed in a manner presented in Algorithm 1. In the first part of the task, the arm DS moves under the influence of the attractor placed at the via-point. The hand DS is driven by the attractor placed at the hand configuration when the palm reaches the closest point (along the trajectory computed ahead of time) to the obstacle. Coupling the hand motion, with respect to the obstacle, is advantageous because it provides a preshape of the hand such that collisions between the fingers and the obstacle are eluded during obstacle avoidance manipulation, even in scenarios where the obstacle is suddenly perturbed during the ongoing task (see Fig. 12). Our approach for adapting the reaching hand motion to avoid obstacles is motivated by several studies that have reported significant effects of the obstacle on all aspects of grasp kinematics (e.g., grip duration, grip aperture, time to peak aperture and distance to peak aperture) (Saling et al. 1998; Tresilian 1998; Mon-Williams et al. 2001). Tresilian (1998) interprets these effects as subtle adjustments of the transport and grip components which support obstacle avoidance. In their obstacle avoidance experiment, Saling et al. (1998) observed a systematic high correlation of arm transport parameters (transport time, time to peak velocity, time to peak acceleration, etc.) with almost all grip kinematic parameters (grip closure time, time to peak aperture, time to peak opening velocity, grip opening velocity, etc.). This result is a very strong indication that the arm and the hand remain coupled even when obstacles cause considerable alternations of prehensile motion, compared with the no-obstacle condition.

The goal hand configuration for passing the obstacle at the closest distance is obtained by observing the average hand configurations of our subjects in obstacle avoidance trials. We adapted, with slight modifications, the computed average hand configuration to match the kinematics of the iCub's hand. We did a similar procedure to obtain the goal hand configurations with respect to the target object.

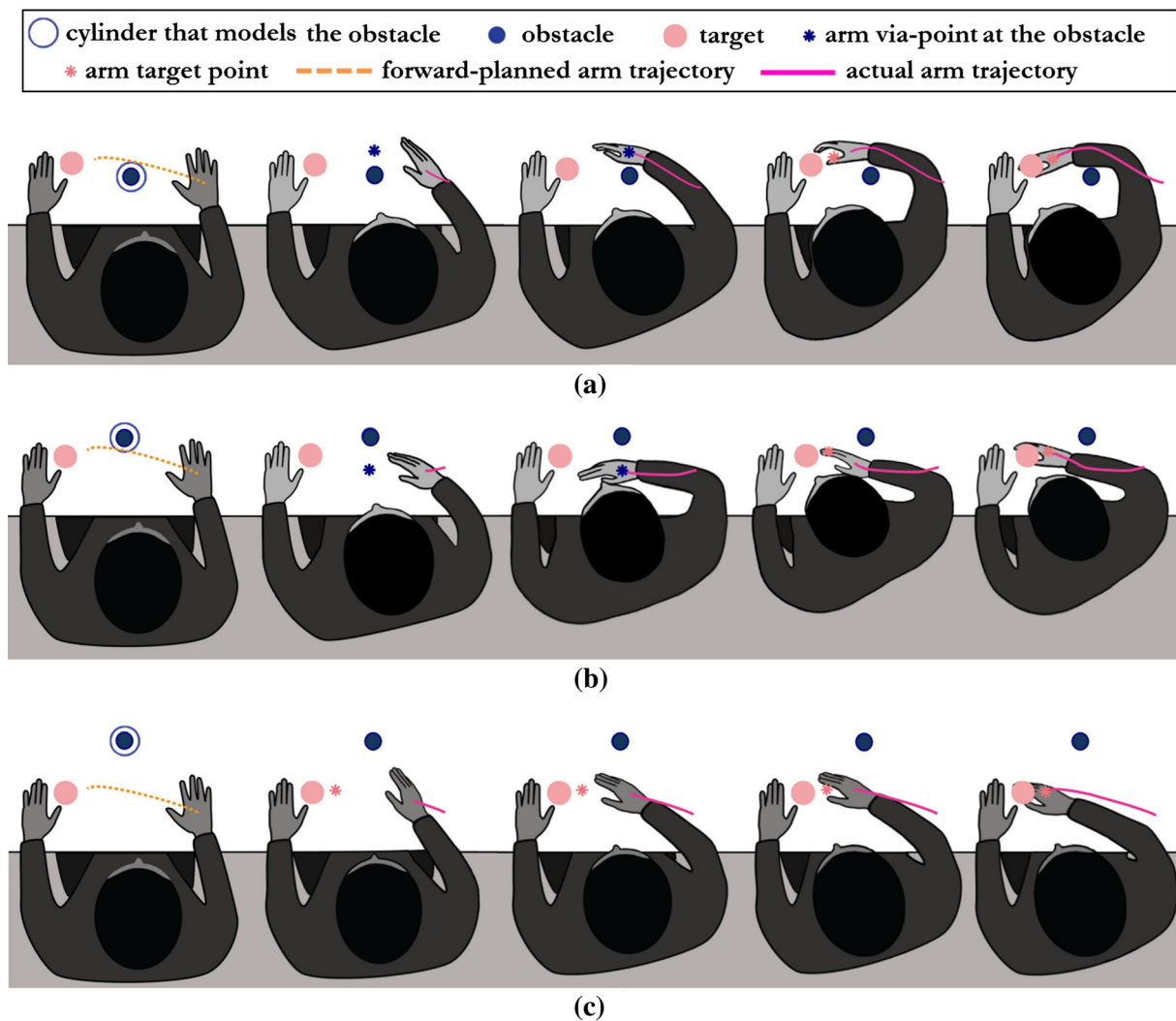
The position of the via-point is determined with respect to the obstacle, such that its displacement vector from the obstacle position is oriented in either an anterior or ventral direction, for the length that corresponds to some safety distance  $d_{safety}$  between the centroid of the palm and the obstacle. We choose the direction of a displacement of the via-point (anterior or ventral) to correspond to a side of the obstacle where a collision is estimated to occur. In the second part of the task, after the obstacle is passed, CDS is driven toward the object to be grasped. As mentioned before, hand adaptation, with respect to the obstacle, serves to support collision avoidance; whereas hand adaptation, with respect to the target, assures coordinated and stable grasping of the target as the arm reaches it. Predefining the safety distance at which the hand passes the obstacle is based on the study of Dean and

Brüwer (1994) who found that participants kept a minimum distance between the pointer and obstacles when performing planar pointing arm movements. In our human study, a measured mean value of this safety distance is 0.142 m with a small value of standard deviation 0.01 m, which can be considered as a consistent observation of the mechanism employed by the motor control system to keep the limb at the safety distance from obstacle, as presented in Dean and Brüwer (1994).

The arm end-effector passing through the via-point at  $d_{safety}$  from the obstacle and hand adaptation, with respect to the obstacle, ensures that the hand will not collide with the obstacle. However, the end-effector obstacle avoidance mechanism, we just described, considers solely collisions with the end-effector and hence ignores collision with the rest of the arm. We benefit from controlling the arm in Cartesian coordinates and from having an efficient inverse kinematics (IK) solver (Pattacini et al. 2010) that is able to handle two tasks: to find suitable joint configuration (primary task) and to keep solutions as close as possible to a desired arm rest posture (secondary task). By having the IK solver that can solve for reaching the Cartesian position by trying to keep joints close to a given rest posture, we can modulate the robot's motion in the operational space by providing joint rest postures suitable for obstacle avoidance. Our approach to the problem of finding suitable joint postures is to learn these joint postures from human demonstrations, as human demonstrations in obstacle avoidance tasks encode inherently favorable joint configurations.

Here we learn correlations between the joints that provide major contributions in obstacle avoidance manipulation and arm position in the operational space. The joints chosen to define the rest position are torso pitch and yaw, and the shoulder joints corresponding to adduction–abduction and flexion–extension. Hence, we proceed with learning the joint probability distribution  $\mathcal{P}(q, x)$ , where  $q \in \mathbb{R}^4$  denotes the joint rest posture and  $x \in \mathbb{R}^3$  denotes the Cartesian position of the arm.

An adaptation of the arm posture for obstacle avoidance is done in the following manner. When reaching for a visuomotor target (the obstacle object or the grasping object), the CDS system infers the state velocities, as explained earlier. By integrating the arm velocity, we obtain a new arm state. By taking the posterior mean estimate of  $\mathcal{P}(q | x)$ , we infer a favorable rest posture. Finally, the IK solver optimizes for joint angles that correspond to the desired Cartesian position, while trying to keep the four joints as close as possible to the suggested values from the model. Figure 11 illustrates our obstacle avoidance scheme. While this does not ensure that the robot's arm will never collide with the obstacle, in practice, we found that this resulted in a successful obstacle avoidance motion.



**Fig. 11** A scheme that illustrates forward planning and obstacle avoidance. After forward integrating the CDS model, an obstacle object (*dark blue disk*) is identified as an obstructing object if the estimated arm motion (*dashed orange line*) intersects with a cylinder (*dark blue circle*) that models the obstacle (certain collision), or when the cylinder lies within the area where it is very likely that it will collide with the forearm (very likely collision). If the obstacle object is identified to obstruct the intended motion, then the motion of the visuomotor system is segmented: from the start to the obstacle and from the obstacle to the target. When reaching to avoid the obstacle, the arm DS moves

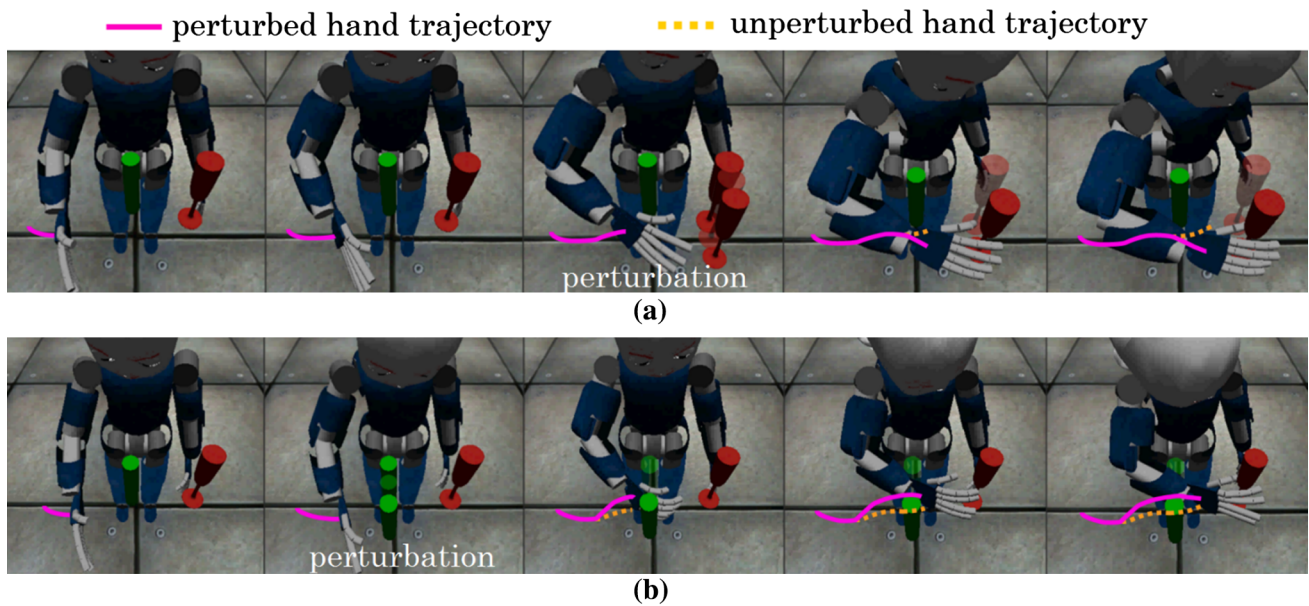
under the influence of the attractor placed at the via-point with respect to the obstacle (*dark blue star*). The direction of a displacement of the via-point (anterior or ventral) is chosen to correspond to a side of the obstacle where a collision is estimated to occur: anterior side (**a**) or ventral side (**b**). If forward planning scheme does not detect collision with the obstacle object (**c**), the visuomotor system is driven to the target object, i.e., the obstacle is ignored. The *light red star* represents the goal arm position with respect to the target object (*light red disk*). Figures show execution of eye–arm–hand coordination from the start of the task (*left*) until the successful grasp completion (*right*) (color figure online)

### 3.4 Robot vision system

The requirements for real-time adaptation to perturbations in dynamic environments impose the demand for real-time update of information obtained from the sensory system. In order to compute the position of objects in every cycle of the control loop, the total time devoted to visual computation in our system has to be reduced to the order of  $\sim 10$  ms for both cameras in the binocular setup of the iCub robot. This is a very hard constraint to achieve in a robotic sys-

tem, even by using modern computing hardware with multi-core processing units. In order to achieve the aforementioned requirement, we designed the visual system to use minimal computational resources.

We use an image processing scheme similar to the one proposed in [Metta et al. \(2004\)](#). We convert  $320 \times 240$  images streamed from the cameras to  $150 \times 150$  log-polar images. By transforming the images to the log-polar domain, we reduce the amount of visual information to be processed, affecting neither the field of view nor the image resolution at the



**Fig. 12** Experiments of visually guided reaching and grasping in the iCub’s simulator, with the presence of the obstacle and with perturbations. The obstacle is an intermediary target for the visuomotor system, hence obstacle avoidance is divided in two sub-tasks: from the start position to the obstacle (via-point) and from the obstacle to the grasping object. Figures show execution of eye–arm–hand coordination from the start of the task (*left*) until the successful grasp completion (*right*). Figures in the *upper row* (a) present a scenario when the target object (red champagne glass) is perturbed during motion (perturbation occurs in the third frame from *left*). Visuomotor coordination when the obstacle

is perturbed during manipulation is shown in the *bottom row* (perturbation in the second frame). The *orange line* shows the trajectory of the hand if there is no perturbation. The *purple line* is the actual trajectory of the hand after perturbation, until successful grasping. In both scenarios (target perturbed and obstacle perturbed), the visuomotor system instantly adapts to the perturbation and drives the motion of the eyes and the arm and the hand to a new position of the object (color figure online)

fixation point. Besides the computational benefits, log-polar mapping is biologically plausible because it approximates the cone distribution in the retina and the mapping from the cone cells to the primary visual cortex of primates (Javier Traver and Bernardino 2010). The image processing is done in the RGB color space, by using a pixel-by-pixel color segmentation algorithm. The same procedure is applied for detection of the target and the obstacle, thus for simplicity of explanation, we will here use the term “object.” After the images are segmented, we apply binary morphological operations to remove outliers, and we group segmented regions in blobs. The centroid of the biggest blob in each image is back-projected from the log-polar domain to the original image coordinates. The distance between the principal point of one camera (we chose the right camera) and the center of the object blob in the visual field represents the eye state  $\xi_e$ , which is the input to the gaze DS. The position estimation of the objects in the workspace is done by triangulating the centroids of the blobs for the left and right camera. The other camera is controlled in a coordinated manner such that both cameras have a fixation point at the estimated head–object distance in the Cartesian coordinates. The distance between the hand and the estimated position of the object represents the arm state

$\xi_a$  that is the input to the arm DS. Algorithm 1 illustrates the flow of visuomotor information processing in our model.

The decreasing visual acuity from the fovea to the periphery implies that we get a more precise estimate of the object position at the point of fixation, and the less accurate estimation in the periphery of the visual field. Because we control the gaze and embed the gaze state to the motor control mechanism, we can inherently and efficiently deal with imprecision in the position estimation associated with non-uniform visual acuity in log-polar images. The CDS drives the gaze, arm and hand toward the object using the pose information (in retinal and Cartesian coordinates) obtained from the vision system. As the gaze moves toward the object in every cycle of the control loop, we update the system with a more precise re-estimate of the object position. Before the hand comes close to the object, the gaze fixates the object, and we get the precise information about the object position, which is crucial for successful grasping and obstacle avoidance. Our time-independent CDS automatically adapts to the re-estimate of the object positions obtained from such non-uniform resolution processing scheme.

For the experiments with the real iCub robot, we use the Viola–Jones detector (Viola and Jones 2001) in addi-

tion to the basic color-based segmentation. We use the additional detector in order to eliminate false-positives detections that are a common consequence of color-based segmentation in an unstructured workspace. In other words, we use this detector to verify our color-based detection. The Viola–Jones detector operates on the images streamed from the camera, not in the log-polar domain. When both detectors agree, we update information about positions of the objects in the workspace, when the detectors do not agree we rely on the previous agreed position. Because the Viola–Jones detector is more computationally demanding, we run it once in every 4 cycles of the control loop.

## 4 Results

### 4.1 Model learning

We learn the CDS model by using the data gathered during the human trials, described in Sect. 2. The parameters of the SEDS algorithm (i.e., maximum number of iterations and optimization criterion) and the number of Gaussian mixtures are determined by using a grid-search with 10-fold cross-validation on the RMSE between the recorded motion and retrieved trajectories from the model. The list of parameter combinations is sorted in ascending order with respect to a value of the RMSE. For each combination of parameters, we visually assess regression plots retrieved from the model. This method is necessary because the small value of the RMSE between the trajectories retrieved from the model and the demonstrated trajectories does not necessarily imply that the inferred paths always have natural-looking and smooth profiles. In other words, the measure of the RMSE provided an initial pool of good candidates, whereas we made the final choice based on the smoothness and the “natural” profile of retrieved paths. The plots for the model we chose are represented in Fig. 10.

We use  $\Psi_e(\xi_e) = \|\cdot\|$ ,  $\Psi_a(\xi_a) = \|\cdot\|$  and the values of parameters  $\alpha_a$ ,  $\alpha_h$ ,  $\beta_a$  and  $\beta_h$  are set to 1. For the choice of the eye–arm coupling function, we tested performances of four different coupling functions: (1.)  $\Psi_e(\xi_e) = \xi_e^2$  (vertical gaze coordinate), (2.)  $\Psi_e(\xi_e) = \xi_e^1$  (horizontal gaze coordinate), (3.)  $\Psi_e(\xi_e) = \xi_e$  (both gaze coordinates) and (4.)  $\Psi_e(\xi_e) = \|\cdot\|$ . We used the average absolute point-to-point differences from all demonstrated trajectories and retrieved trajectories from the models as a measure of how well these coupling functions perform. The best results are obtained by the norm coupling function. Our motivation for using  $\|\cdot\|$  function for arm–hand coupling is based on our previous work in hand–arm coupling, see Shukla and Ballard (2011). Our choice of these particular coupling functions can be considered biologically plausible. The choice of  $\|\cdot\|$  for arm–hand coupling is supported by the physiolog-

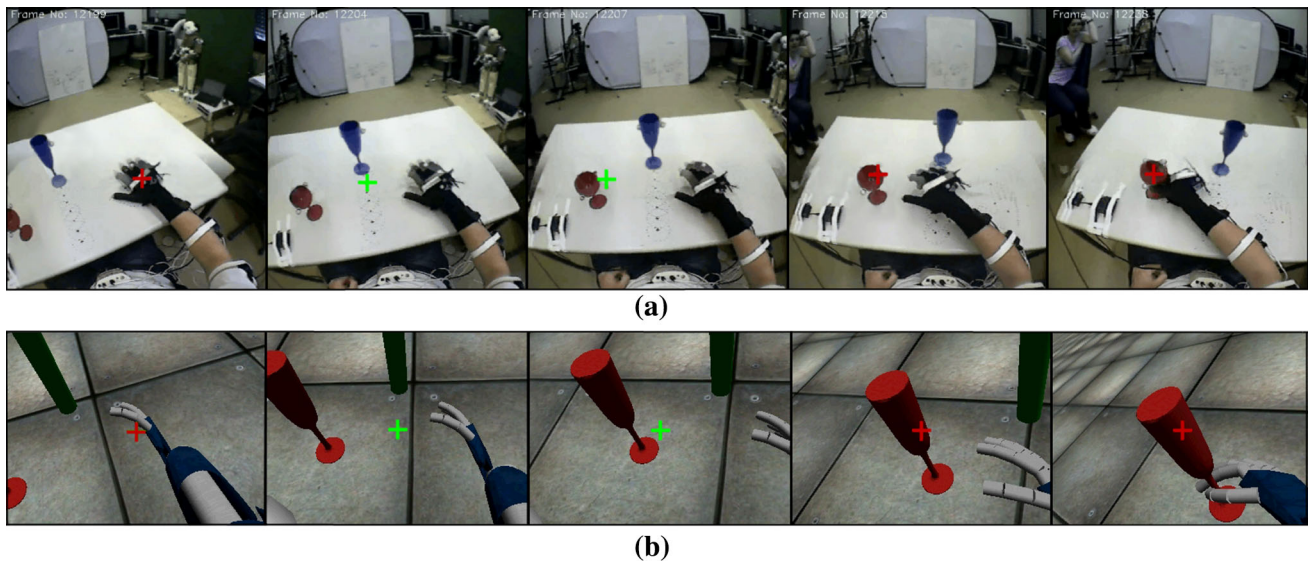
ical studies (Haggard and Wing 1991, 1995) that reported strong coupling of the hand preshape with respect to the distance from the target object in reach-for-grasping tasks. The choice of  $\|\cdot\|$  for eye–arm coupling function is supported by the fact that retinal distance in foveated vision directly affects the quality of visual information that is used by the motor system for planning and performing manipulation, as visual acuity decreases with distance from the fovea (Land et al. 1999; Land 1999; Liversedge and Findlay 2000; Hayhoe and Ballard 2005). All  $\alpha$  and  $\beta$  parameters are set to 1 in order to ensure an unaltered reproduction profile of visuomotor coordination learned from recorded human demonstrations.

### 4.2 Model validation for robot control

We conduct a set of experiments with the iCub robot to evaluate the performances of our approach for the visuomotor coordination. Due to hardware constraints of the real robot, we perform perturbation experiments and experiment with obstacle avoidance in the iCub simulator. Unperturbed obstacle-free reaching and grasping experiments are conducted with the real robot.

In our experiments, we validate the ability of the CDS controller on the iCub robot to reproduce the same task of visually guided obstacle-free reaching and grasping similar to the one that humans performed in our trials, together with the advocated robustness of the model to perturbations and the ability to handle the obstacles in the workspace.

We present here the most demanding experiment we perform to validate our approach. In each run, the object to be grasped is placed at a randomly computed position within a 15 cm cube in the workspace. Figure 12 shows an obstacle scenario where we test coordinated manipulation with sudden perturbations of the target object and the obstacle, respectively. To introduce perturbations on-the-fly during reaching for the target, we implement a pre-programmed routine in the simulator to abruptly change position of the object (target or obstacle) when the hand approaches it at some predefined distance, which varies from trial to trial from 0.09 m to 0.15 m. The robot’s end-effector avoids the obstacle when reaching for grasping in two task segments: (1) start position  $\rightarrow$  via-point at  $d_{safety}$  from obstacle and (2) via-point at  $d_{safety}$  from obstacle  $\rightarrow$  grasping object. This safety distance in the human trials is  $d_{safety} = 0.142 \pm 0.01$  m. We rescale the safety distance from human trials by 2, because the dimensions of the iCub are similar to those of a 3.5-year-old child; hence, it has a smaller workspace than our adult subjects. Once the obstacle is reached, the target for the visuomotor system is changed, and the eye–arm–hand motion is directed to the object to be grasped. The IK solver adapts the arm rest posture to be as close as possible to the output inferred from the model learned from human demonstrations. Figure 13



**Fig. 13** A comparison of human visuomotor coordination and visuomotor behavior of the real robot. The visuomotor coordination profile the robot produces **(b)** is highly similar to the pattern of coordination

that was observed in the human trials **(a)**. The figures from *left to right* show snapshots of the execution of eye–arm–hand coordination from the start of the task (*left*) until the successful grasp completion (*right*)

shows how the human subjects ignore the obstacle when it does not obstruct the intended motion, and the same pattern produced by our visuomotor robotic controller.

Because the eye state is the distance between the position of gaze and the position of a visual target in retinal coordinates, and the arm state is represented with respect to the position of the object in the Cartesian space, both variables are instantly updated when the perturbation occurs. The DS of the eyes adapts independently to the perturbation. The behavior of the DS of the arm is modulated via the eye–arm coupling function, and the hand DS is modulated via the arm–hand coupling. Such modulation ensures that the learned profile of eye–arm–hand coordination will be preserved and that the hand will re-open as the object is perturbed away from it, see Fig. 12. Besides the anthropomorphic profile of visuomotor coordination (Fig. 14), the gaze–arm lag allows for enough time to foveate at the object, to re-estimate object’s pose and to compute suitable grasp configuration for the hand before it approaches too close to the object.

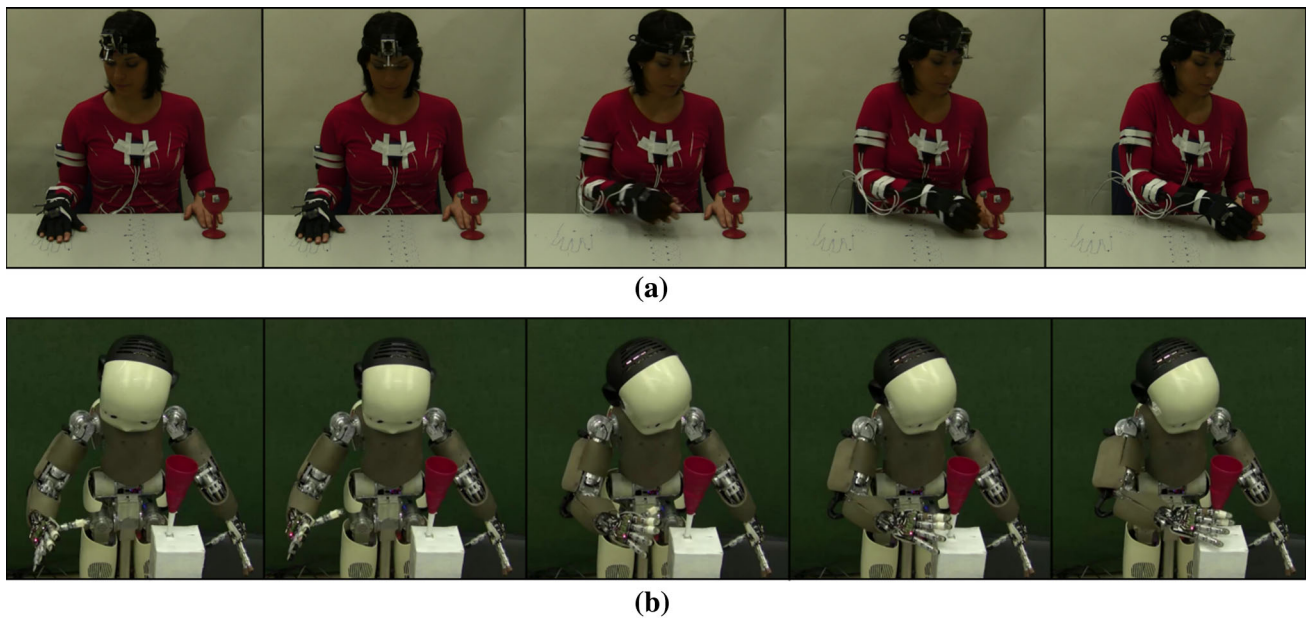
In setups where the arrangement of the obstacle and target differs to a moderate extent compared the setup used in the human demonstrations, the robot successfully grasps the target object, in both obstacle avoidance and no-obstacle tasks, as shown in the experiments presented in the paper and in the accompanying online video. Scene setups that are significantly different often imply a substantially different approach of the hand to the target object than the one seen in the demonstrations. In our case, this occasionally results either in collision of the fingers with the object prior to grasping or incomplete closure of the fingers on the target object.

This is not due to our gaze–arm–hand controller, but rather is due to the fact that we rely on a predefined set of the final hand configurations obtained from human trials. With moderate changes to how the hand approaches an object with complex geometry, like the champagne glass in our experiment, the set of stable hand configurations sometimes can change significantly. In order to increase the rate of grasping in scenarios that substantially differ from the setup in the demonstrations, we would need to use one of the robotic grasp synthesis algorithms to generate the final hand configuration (Sahbani et al. 2012).

The experiments presented here, with several additional experiments on the real robot and in the simulator, are available online at <http://lasa.epfl.ch/videos/downloads/LukicBiologicalCybernetics2012.mp4>.

### 5 Discussion

Our approach to the problem of controlling robotic eye–arm–hand coordination takes inspiration in the pattern of visuomotor coordination displayed by humans. The CDS control framework drives the gaze, the arm and the hand in a synchronous manner. This approach harvests the major benefits of encoding motion with time-invariant DS: robustness to spatio-temporal perturbations and instant re-planning of motion when perturbations occur. The CDS global stability guarantees that the eye, the arm and the hand will reach the target in retinal, operational and grasp space, respectively, even when spatial and temporal perturbations are present.



**Fig. 14** The visuomotor system ignores an obstacle object when it is not relevant for manipulation, i.e., the obstacle object that does not affect intended motion is not visually salient for the gaze. Analysis of the WearCam recordings from the human trials (a) reveals that subjects do not fixate the obstacle object (*blue champagne glass*) in the workspace when it does not obstruct intended reaching and grasping

movement. Our CDS eye–arm–hand model shows the same behavior (b), ignoring the obstacle object (*green cylinder*), when the forward planning scheme estimates that the object does not obstruct prehensile movement. Snapshots show task from the start (*left*) until completion of the successful grasp (*right*) (color figure online)

### 5.1 Discussion on controller architecture

Our approach uses human demonstrations, which provide a model to guide the dynamics of motion as in open-loop visuomotor transformation techniques (Hoffmann et al. 2005; Natale et al. 2005, 2007; Hulse et al. 2009). A stable model of the high-dimensional visuomotor coordination can be learned by using only several human demonstrations, making it a very efficient, fast and intuitive way to estimate parameters of a robot visuomotor controller. The generalization abilities of the CDS framework ensure coordinated behavior of the visuomotor controller, even when the motion is abruptly perturbed outside the region of the provided human demonstrations. Similar to visual servoing (Espiau et al. 1992; Mansard et al. 2006; Natale et al. 2007; Chaumette and Hutchinson 2008), it performs a closed-loop control, and hence, it ensures that the target can be reached under perturbations. Our approach inherently combines learned visuomotor transformations and visual servoing characteristics, thus it eliminates the need to rely on an external ad hoc mechanism for switching between the two modes (Natale et al. 2007). Coupling profiles for eye–arm and arm–hand systems can be modulated, thus allowing us to adjust the behavior of each slave system with respect to control signals flowing from the corresponding master system. Our eye–arm–hand controller drives the hand motion in synchronization with the

gaze and arm motion. This provides a means to build a compact model of the visuomotor coordination, in a biologically inspired manner, without pre-programming the hand control policy. The major building blocks that constitute the architecture of our controller are the gaze DS, the arm DS and the hand DS. These blocks are coordinated by using the gaze–arm and the arm–hand coupling functions. Each coupling function transfers the information about the state of a master controller to signals that modulate the behavior of a slave controller. The gaze controller is the master controller of the arm, and the arm controller is the master of the hand. This control architecture is supported with the existing evidence of gaze leading arm motion (Abrams et al. 1990; Johansson et al. 2001; Hayhoe et al. 2003) and the existing reports on coupling between the transport and the grip component in the studies of prehensile motion (Haggard and Wing 1991, 1995).

### 5.2 Discussion on obstacle avoidance

We hypothesized that the visuomotor system treats the obstacle as an intermediary target. Evidence of a systematic pattern that the gaze precedes and leads the motion of the arm through the different landmarks, defining the stages of a sequential task, supports this hypothesis (Johansson et al. 2001).

We extended our original framework CDS (Shukla and Billard 2011) for visuomotor coordination on obstacle avoidance such that the task is executed in two segments: from the start to the obstacle and from the obstacle to the target.

We took inspiration in the forward planning mechanisms (Wolpert et al. 1998, 2001) in the design of our approach for estimating the consequences of planned arm movements, specifically to detect objects that obstruct the intended reach-for-grasp actions and to identify them as obstacles.

In our obstacle avoidance mechanism, the gaze is as a constituting element of the overall visuomotor mechanism, and it is actively controlled and intermingled with manipulation requirements and plans.

During obstacle avoidance, the primary modulation of the arm is controlled in the operational space, which, together with controlled hand preshape, ensures that the end-effector avoids the obstacle. The rest postures suitable for obstacle avoidance are provided to the IK solver. We learn these rest postures from the data gathered when the subjects avoid the obstacle in prehensile motion.

It is important to mention that our obstacle avoidance scheme does not have the full strengths of methods such as rapidly exploring random trees (RRTs) (Kuffner and LaValle 2000) for reaching in very complex workspaces, but it endows the visuomotor system with instant reactions to perturbations, thus providing means for the rapid handling of a relatively simple obstacle in the workspace.

### 5.3 Future work

In our controller, the flow of control signals is monodirectional and it is oriented in the direction eyes  $\rightarrow$  arm  $\rightarrow$  hand. However, some studies report that the control signals also flow from the hand to the eyes (Fisk and Goodale 1985; Neggers and Bekkering 2000), and from the hand to the arm (Timmann et al. 1996). Hence, it could be worth addressing the potential benefits of modeling and using bidirectional visuomotor control schemes in robotics. Having the control signals flow in the opposite direction, hand  $\rightarrow$  arm  $\rightarrow$  eyes, is useful, for instance, to trigger a reactive motion of the gaze and the arm when facing an unexpected displacement of the hand, such as when the hand inadvertently touches an obstacle.

We present a model of coupling between the gaze movements and the arm–hand motion. Although overt movements of the eyes leading the arm represent a dominant visuomotor pattern in natural tasks, several physiological studies have shown that humans can perform pointing and grasping tasks toward extrafoveal targets with the eye movements suppressed (Prablanc et al. 1979; Abrams et al. 1990). The physiological studies and corresponding robotic models (Schenck et al. 2011) have reported significantly decreased success in grasping extrafoveal targets compared with grasp-

ing foveated targets. In the study of Johansson et al. (2001) where eye movements were prevented during manipulation with obstacles, the overall performances degraded, which was observed in frequent collisions with the obstacle. The reduced performances associated with grasping extrafoveal targets are not desirable for robotic applications, where the maximal efficiency of tasks is one of the primary goals. However, we believe that modeling of grasping of extrafoveal targets is an important issue to address in order to have more biologically plausible computational models. The main challenge in using the programming-by-demonstration framework for modeling coupling between covert attentional spotlight shifts and the arm movements is the methods of measuring allocation of covert visual attention during manipulation tasks. Measuring covert attentional shifts can be achieved either by analyzing modulation of microsaccades for mapping them to shifts of covert attention (Engbert et al. 2003) or by relying on a variant of a secondary discrimination task, which is a more common technique (Baldauf and Deubel 2010). These methods can indicate the position of the covert attention spotlight in the visual field with a rather coarse resolution, which at this moment represents a major obstacle for achieving this goal. Current work in our group is investigating the possibility to embed covert attention in the mechanism of visuomotor coordination.

In our robotic implementation, we select the centroid of the object (obstacle and target) as the fixation point for the gaze. However, this simplified scheme of selecting the fixation points at the object might be upgraded in order to improve both biological plausibility and the computational benefits of using active vision. From physiological studies, it is known that the gaze fixations are driven to regions of the target contact points in grasping, whereas in viewing tasks, the gaze is directed to the object's centroid (Brouwer et al. 2009). An explanation for this result is that fixations during grasping are focused on the object's contact parts because the eyes provide visual feedback for motor control of the fingers in grasping scenarios. These contact parts are mostly close to the boundary of an object. The gaze is more likely to fall on the edges of obstacles, in both manipulation tasks (Johansson et al. 2001) and in navigation (Rothkopf and Ballard 2009), which can be explained by taking visual information for path planning for obstacle avoidance. We observed the same effects in our human trials. However, at this point, there are not yet computational models that tackle problems of selecting optimal fixation points at the target object and obstacles. We believe that it would be tremendously useful to tackle this scientific problem. Recent work on active segmentation might offer the computational ground for tackling these problems (Mishra et al. 2009a,b).

Furthermore, in this work, we assumed a constant value for the safety margin  $d_{safety}$  between the arm's via-point and the obstacle. In our robot experiments, we used rescaled

value of the safety distance measured from human trials in experiment 1. The results of experiment 2, when the obstacle is moved along the midline of the desk, indicated that this safety distance was kept quasi-constant across subjects, and for all trials where the hand would have touched the obstacle if moving with the regular pattern of motion. However, there is no reason to think that this safety margin is a constant, pre-set factor. Some studies showed that this safety margin was modulated by the speed of movement (e.g., faster prehensile movements are associated with a greater safety distance) (Tresilian 1998; Mon-Williams et al. 2001) and “a variety of psychological factors related to the cost that a person attaches to a collision” (Tresilian 1998). It would be of great importance, both for motor control science and robotic obstacle avoidance applications, to model this safety distance, rather than to consider it as a preset factor (Bendahan and Gorce 2006). One approach to model this safety margin is to estimate it from the data recorded from human demonstrations by varying task conditions across trials (e.g., shape and size of an obstacle, relative positions of objects in the workspace, required speed of manipulation, task and objectives) by using machine learning techniques.

Robotic engineers have studied avoidance of multiple obstacles for a long time (Khatib 1986; Lumelsky and Skewis 1990; Simmons 1996; Kavraki et al. 1996; Kuffner and LaValle 2000), but it is rather surprising that only a small number of studies in motor control, physiology and visual science studied human manipulation in tasks where several obstacles occupy the workspace. In their study, Mon-Williams et al. (2001) reported on the greater effect of two obstacles on the movement time, maximum grip aperture and peak speed compared with the one-obstacle case. Rothkopf and Ballard (2009), who studied human navigation in an immersed graphic environment, reported that subjects fixate the edges of obstacles for the purpose of planning a walking path for obstacle avoidance. Aivar et al. (2008) provided evidence that fast arm responses to the displacement of obstacles are triggered by a reaction to retinal motion of moving obstacles. Many important questions still remain unanswered. Do humans assess multiple obstacles in a sequential manner, assigning priorities to obstacles according to an estimated risk of collision, or simultaneously? How are the eyes, the arm and the hand coordinated when handling multiple obstacles in reaching and grasping tasks? How does the human visuomotor and planning systems react when one or several obstacles are perturbed in the workspace during prehensile tasks? Studying visuomotor coordination in natural prehensile tasks with several non-target objects in the workspace could provide more insights into these questions.

## 5.4 Conclusion

Our human study contributes a quantitative assessment of the eye–arm coordination when performing obstacle avoidance, an issue which has received little attention to date. Precisely, it demonstrates that obstacle avoidance is included in forward planning and modulates the coordinated pattern of eye–arm motion in a distinctive way. The results of the study: (a) quantify the phase relationship between these gaze and the arm systems, so as to inform robotic models; and (b) provide insights how the presence of an obstacle modulates this pattern of correlations. We show that the notion of workspace travelled by the hand is embedded explicitly in a forward planning scheme that allows subjects to determine when and when not to pay attention to the obstacle. Importantly, to complement these observations, we provide a computational model of both the eye–arm–hand coupling and the modulation of the obstacle in this forward planning scheme.

**Acknowledgments** This work was supported in part by EU projects POETICON++ (FP7-ICT-288382) and FIRST-MM (FP7-ICT 248258) and Fundação para a Ciência e a Tecnologia (FCT) doctoral grant (SFRH/BD/51072/2010) under IST-EPFL Joint Doctoral Initiative.

## References

- Abrams R, Meyer D, Kornblum S (1990) Eye-hand coordination: oculomotor control in rapid aimed limb movements. *J Exp Psychol Hum Percept Perform* 16(2):248
- Aivar M, Brenner E, Smeets J (2008) Avoiding moving obstacles. *Exp Brain Res* 190(3):251–264
- Alberts JL, Saling M, Stelmach GE (2002) Alterations in transport path differentially affect temporal and spatial movement parameters. *Exper Brain Res* 143(4):417–425
- Aloimonos J, Weiss I, Bandyopadhyay A (1988) Active vision. *Int J Comput Vis* 1(4):333–356
- Andersen RA, Cui H (2009) Intention, action planning, and decision making in parietal–frontal circuits. *Neuron* 63(5):568–583
- Bajcsy R (1988) Active perception. *Proc IEEE* 76(8):966–1005
- Bajcsy R, Campos M (1992) Active and exploratory perception. *CVGIP Image Underst* 56(1):31–40
- Baldauf D, Deubel H (2010) Attentional landscapes in reaching and grasping. *Vis Res* 50(11):999–1013
- Ballard D (1991) Animate vision. *Artif Intell* 48(1):57–86
- Ballard DH, Hayhoe MM, Pelz JB (1995) Memory representations in natural tasks. *J Cogn Neurosci* 7(1):66–80
- Bendahan P, Gorce P (2006) A neural network architecture to learn arm motion planning in grasping tasks with obstacle avoidance. *Robotica* 24(2):197–204
- Berthier NE, Clifton RK, Gullapalli V, McCall DD, Robin DJ (1996) Visual information and object size in the control of reaching. *J Mot Behav* 28(3):187–197
- Bishop C (2007) Pattern recognition and machine learning (information science and statistics). *Pattern Recognit* 4(2):1–748



- Bowman M, Johansson R, Flanagan J (2009) Eye-hand coordination in a sequential target contact task. *Exp Brain Res* 195(2):273–283
- Brouwer A, Franz V, Gegenfurtner K (2009) Differences in fixations between grasping and viewing objects. *J Vis* 9(1):1–8
- Castiello U, Bennett K, Mucignat C (1983) The reach to grasp movement of blind subjects. *Exp Brain Res* 96(1):152–162
- Castiello U, Bennett K, Stelmach G (1993) Reach to grasp: the natural response to perturbation of object size. *Exp Brain Res* 94(1):163–178
- Chaumette F, Hutchinson S (2008) Visual servoing and visual tracking. In: Siciliano B, Khatib O (eds) *Springer Handbook of Robotics*. Springer, Berlin, Heidelberg, pp 563–583
- Dalton K, Nacewicz B, Johnstone T, Schaefer H, Gernsbacher M, Goldsmith H, Alexander A, Davidson R (2005) Gaze fixation and the neural circuitry of face processing in autism. *Nat Neurosci* 8(4):519–526
- Dean J, Brüwer M (1994) Control of human arm movements in two dimensions: paths and joint control in avoiding simple linear obstacles. *Exp Brain Res* 97(3):497–514
- Deubel H, O'Regan JK, Radach R (2000) Attention, information processing, and eye movement control. In: Kennedy A, Radach R, Heller D, Pynte J (eds) *Reading as a perceptual process*. Elsevier, Oxford, pp 355–374
- Engbert R, Kliegl R et al (2003) Microsaccades uncover the orientation of covert attention. *Vis Res* 43(9):1035–1045
- Espiau B, Chaumette F, Rives P (1992) A new approach to visual servoing in robotics. *IEEE Trans Robot Autom* 8(3):313–326
- Fisk J, Goodale M (1985) The organization of eye and limb movements during unrestricted reaching to targets in contralateral and ipsilateral visual space. *Exp Brain Res* 60(1):159–178
- Gentilucci M, Toni I, Chieffi S, Pavesi G (1994) The role of proprioception in the control of prehension movements: a kinematic study in a peripherally deafferented patient and in normal subjects. *Exp Brain Res* 99(3):483–500
- Gibson JJ (1950) *The perception of the visual world*. Houghton Mifflin, Boston
- González-Alvarez C, Subramanian A, Pardhan S (2007) Reaching and grasping with restricted peripheral vision. *Ophthalmic Physiol Opt* 27(3):265–274
- Goodale MA (2011) Transforming vision into action. *Vis Res* 51(13):1567–1587
- Goodale MA, Haffenden A (1998) Frames of reference for perception and action in the human visual system. *Neurosci Biobehav Rev* 22(2):161–172
- Grasso R, Prévost P, Ivanenko Y, Berthoz A et al (1998) Eye-head coordination for the steering of locomotion in humans: an anticipatory synergy. *Neurosci Lett* 253(2):115–118
- Haggard P, Wing A (1991) Remote responses to perturbation in human prehension. *Neurosci Lett* 122(1):103–108
- Haggard P, Wing A (1995) Coordinated responses following mechanical perturbation of the arm during prehension. *Exp Brain Res* 102(3):483–494
- Hayhoe M, Ballard D (2005) Eye movements in natural behavior. *Trends Cogn Sci* 9(4):188–194
- Hayhoe M, Shrivastava A, Mruczek R, Pelz J (2003) Visual memory and motor planning in a natural task. *J Vis* 3(1):49–63
- Henderson JM, Hollingworth A (1999) The role of fixation position in detecting scene changes across saccades. *Psychol Sci* 10(5):438–443
- Hesse C, Deubel H (2010) Effects of altered transport paths and intermediate movement goals on human grasp kinematics. *Exp Brain Res* 201(1):93–109
- Hesse C, Deubel H (2011) Efficient grasping requires attentional resources. *Vis Res* 51(11):1223–1231
- Hicheur H, Berthoz A (2005) How do humans turn? head and body movements for the steering of locomotion. In: *IEEE-RAS international conference on humanoid robots (Humanoids)*, IEEE, pp 265–270
- Hoffmann H, Schenck W, Möller R (2005) Learning visuomotor transformations for gaze-control and grasping. *Biol Cybern* 93(2):119–130
- Hulse M, McBrid S, Lee M (2009) Robotic hand-eye coordination without global reference: a biologically inspired learning scheme. In: *IEEE international conference on development and Learning (ICDL)*, IEEE, pp 1–6
- Inhoff AW, Radach R (1998) Definition and computation of oculomotor measures in the study of cognitive processes. In: Underwood G (ed) *Eye guidance in reading and scene perception*. Elsevier, Amsterdam, pp 29–53
- Jacob R, Karn K (2003) Eye tracking in human-computer interaction and usability research: ready to deliver the promises. *Mind* 2(3):4
- Jakobson L, Goodale M (1991) Factors affecting higher-order movement planning: a kinematic analysis of human prehension. *Exp Brain Res* 86(1):199–208
- Jamone L, Natale L, Nori F, Metta G, Sandini G (2012) Autonomous online learning of reaching behavior in a humanoid robot. *Int J Humanoid Robot* 9(03):1–26
- Javier Traver V, Bernardino A (2010) A review of log-polar imaging for visual perception in robotics. *Robot Auton Syst* 58(4):378–398
- Jeannerod M (1984) The timing of natural prehension movements. *J Mot Behav* 16(3):235–254
- Johansson R, Westling G, Bäckström A, Flanagan J (2001) Eye-hand coordination in object manipulation. *J Neurosci* 21(17):6917–6932
- Johansson RS, Flanagan JR, Johansson RS (2009) Sensory control of object manipulation. *Sensorimotor control of grasping: physiology and pathophysiology*. Cambridge University Press, Cambridge
- Kavraki LE, Svestka P, Latombe JC, Overmars MH (1996) Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Trans Robot Autom* 12(4):566–580
- Khansari-Zadeh S, Billard A (2011) Learning stable nonlinear dynamical systems with Gaussian mixture models. *IEEE Trans Robot* 27(5):943–957
- Khansari-Zadeh SM, Billard A (2012) A dynamical system approach to realtime obstacle avoidance. *Auton Robots* 32(4):433–454
- Khatib O (1986) Real-time obstacle avoidance for manipulators and mobile robots. *Int J Robot Res* 5(1):90–98
- Kuffner Jr J, LaValle S (2000) Rrt-connect: an efficient approach to single-query path planning. In: *IEEE international conference on robotics and automation (ICRA)*, IEEE, vol 2, pp 995–1001
- Land M (1999) Motion and vision: why animals move their eyes. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 185(4):341–352
- Land M, Mennie N, Rusted J et al (1999) The roles of vision and eye movements in the control of activities of daily living. *Perception* 28(11):1311–1328
- Land MF, Furneaux S (1997) The knowledge base of the oculomotor system. *Philos Trans R Soc Lond Ser B Biol Sci* 352(1358):1231–1239
- Liversedge S, Findlay J (2000) Saccadic eye movements and cognition. *Trends Cogn Sci* 4(1):6–14
- Lukic L, Santos-Victor J, Billard A (2012) Learning coupled dynamical systems from human demonstration for robotic eye-arm-hand coordination. In: *Proceedings of the IEEE-RAS international conference on humanoid robots (Humanoids)*, Osaka, Japan
- Lumelsky V, Skewis T (1990) Incorporating range sensing in the robot navigation function. *IEEE Trans Syst Man Cybern* 20(5):1058–1069
- Mansard N, Lopes M, Santos-Victor J, Chaumette F (2006) Jacobian learning methods for tasks sequencing in visual servoing. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, pp 4284–4290
- Metta G, Gasteratos A, Sandini G (2004) Learning to track colored objects with log-polar vision. *Mechatronics* 14(9):989–1006

- Metta G, Natale L, Nori F, Sandini G, Vernon D, Fadiga L, Von Hofsten C, Rosander K, Lopes M, Santos-Victor J et al (2010) The icub humanoid robot: an open-systems platform for research in cognitive development. *Neural Netw* 23(8–9):1125–1134
- Mishra A, Aloimonos Y, Fah CL (2009a) Active segmentation with fixation. In: 12th international conference on computer vision (ICCV), IEEE, pp 468–475
- Mishra A, Aloimonos Y, Fermuller C (2009b) Active segmentation for robotics. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, pp 3133–3139
- Mon-Williams M, Tresilian J, Coppard V, Carson R (2001) The effect of obstacle position on reach-to-grasp movements. *Exp Brain Res* 137(3):497–501
- Natale L, Metta G, Sandini G (2005) A developmental approach to grasping. In: Developmental robotics AAAI spring symposium, vol 44
- Natale L, Nori F, Sandini G, Metta G (2007) Learning precise 3d reaching in a humanoid robot. In: IEEE international conference on development and learning (ICDL), IEEE, pp 324–329
- Neggers S, Bekkering H (2000) Ocular gaze is anchored to the target of an ongoing pointing movement. *J Neurophysiol* 83(2):639–651
- Noris B, Keller J, Billard A (2010) A wearable gaze tracking system for children in unconstrained environments. *Comput Vis Image Underst* 115(4):476–486
- Paillard J (1982) The contribution of peripheral and central vision to visually guided reaching. In: Ingle D, Goodale M, Marsfield R (eds) *Analysis of visual behavior*. MIT Press, Cambridge, pp 367–385
- Pattacini U, Nori F, Natale L, Metta G, Sandini G (2010) An experimental evaluation of a novel minimum-jerk Cartesian controller for humanoid robots. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, pp 1668–1674
- Paulignan Y, MacKenzie C, Marteniuk R, Jeannerod M (1991) Selective perturbation of visual input during prehension movements. *Exp Brain Res* 83(3):502–512
- Pelissod D, Prablanc C, Goodale M, Jeannerod M (1986) Visual control of reaching movements without vision of the limb. *Exp Brain Res* 62(2):303–311
- Prablanc C, Echallier J, Komilis E, Jeannerod M (1979) Optimal response of eye and hand motor systems in pointing at a visual target. *Biol Cybern* 35(2):113–124
- Purdy KA, Lederman SJ, Klatzky RL (1999) Manipulation with no or partial vision. *J Exp Psychol Hum Percept Perform* 25(3):755
- Rayner K (1998) Eye movements in reading and information processing: 20 years of research. *Psychol Bull* 124(3):372
- Rizzolatti G, Fogassi L, Gallese V (1997) Parietal cortex: from sight to action. *Curr Opin Neurobiol* 7(4):562–567
- Rossetti Y, Stelmach G, Desmurget M, Prablanc C, Jeannerod M (1994) The effect of viewing the static hand prior to movement onset on pointing kinematics and variability. *Exp Brain Res* 101(2):323–330
- Rothkopf C, Ballard D (2009) Image statistics at the point of gaze during human navigation. *Vis Neurosci* 26(01):81–92
- Rothkopf C, Ballard D, Hayhoe M (2007) Task and context determine where you look. *J Vis* 7(14):1–16
- Sahbani A, El-Khoury S, Bidaud P (2012) An overview of 3d object grasp synthesis algorithms. *Robot Auton Syst* 60(3):326–336
- Saling M, Alberts J, Stelmach G, Bloedel J (1998) Reach-to-grasp movements during obstacle avoidance. *Exp Brain Res* 118(2):251–258
- Schenck W, Hoffmann H, Möller R (2011) Grasping of extrafoveal targets: a robotic model. *New Ideas Psychol* 29(3):235–259
- Seara JF, Strobl KH, Schmidt G (2003) Path-dependent gaze control for obstacle avoidance in vision guided humanoid walking. In: IEEE international conference on robotics and automation (ICRA), IEEE, vol 1, pp 887–892
- Shukla A, Billard A (2011) Coupled dynamical system based arm–hand grasping model for learning fast adaptation strategies. *Robot Auton Syst* 60(3):424–440
- Simmons R (1996) The curvature–velocity method for local obstacle avoidance. In: IEEE international conference on robotics and automation (ICRA), IEEE, vol 4, pp 3375–3382
- Sivak B, MacKenzie CL (1990) Integration of visual information and motor output in reaching and grasping: the contributions of peripheral and central vision. *Neuropsychologia* 28(10):1095–1116
- Spijkers WA, Lochner P (1994) Partial visual feedback and spatial end-point accuracy of discrete aiming movements. *J Mot Behav* 26(3):283–295
- Srinivasa SS, Berenson D, Cakmak M, Collet A, Dogar MR, Dragan AD, Knepper RA, Niemueller T, Strabala K et al (2012) Herb 2.0: lessons learned from developing a mobile manipulator for the home. *Proc IEEE* 100(8):2410–2428
- Sung HG (2004) Gaussian mixture regression and classification. PhD thesis, Rice University
- Tatler BW, Hayhoe MM, Land MF, Ballard DH (2011) Eye guidance in natural vision: reinterpreting salience. *J Vis* 11(5):1–23
- Timmann D, Stelmach G, Bloedel J (1996) Grasping component alterations and limb transport. *Exp Brain Res* 108(3):486–492
- Tresilian J (1998) Attention in action or obstruction of movement? A kinematic analysis of avoidance behavior in prehension. *Exp Brain Res* 120(3):352–368
- Triesch J, Ballard DH, Hayhoe MM, Sullivan BT (2003) What you see is what you need. *J Vis* 3(1):86–94
- Vernon D, Hofsten C, Fadiga L (2010) A roadmap for cognitive development in humanoid robots, vol 11. Springer, Berlin
- Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society conference on computer vision and pattern recognition, IEEE, vol 1, p 1–511
- Wolpert D, Miall R, Kawato M (1998) Internal models in the cerebellum. *Trends Cogn Sci* 2(9):338–347
- Wolpert D, Flanagan J et al (2001) Motor prediction. *Curr Biol* 11(18):729