Int J Soc Robot (2011) 3:223–231
DOI 10.1007/s12369-010-0078-3

How Humans Optimize Their Interaction with the Environment: The Impact of Action Context on Human Perception

Agnieszka Wykowska · Alexis Maldonado · Michael Beetz · Anna Schubö

Accepted: 6 September 2010 / Published online: 30 September 2010

Abstract This paper reports empirical findings on human performance in an experiment comprising a perceptual task and a motor task. Such findings should be considered in design of robots, since drawing inspiration from natural solutions not only should prove beneficial for artificial systems but also human-robot interaction should then become more efficient and safe. Humans have developed various mechanisms to optimize the way actions are performed and the effects they induce. Optimization of action planning (e.g., grasping, reaching or lifting objects) requires efficient selection of action-relevant features. Selection might also depend on the environmental context in which an action takes place. The present study investigated how action context influences perceptual processing in action planning. The experimental paradigm comprised two independent tasks: (1) a perceptual visual search task and (2) a grasping or a pointing movement. Reaction times in the visual search task were measured as a function of the movement type (grasping vs. pointing) and context complexity (context varying along one dimension vs. context varying along two dimensions). Results showed that action context influenced reaction times, which suggests a close bidirectional link between action and perception as well as an impact of environmental action context on perceptual selection in the course of action planning. These findings are discussed in the context of application for robotics and design of users' interfaces.

Keywords Action context · Visual perception · Action-perception links

A. Wykowska (✉) · A. Schubö
Department of Experimental Psychology, Ludwig Maximilians Universität, München, Germany
e-mail: agnieszka.wykowska@psy.lmu.de

A. Schubö
e-mail: anna.schuboe@LMU.de

A. Maldonado · M. Beetz
Computer Science Department, Chair IX, Technische Universität, München, Germany

A. Maldonado
e-mail: maldonado@tum.de

M. Beetz
e-mail: beetz@in.tum.de

1 Introduction

When humans perform a particular action such as, for example, reaching for a cup in a cupboard they need not only to specify movement parameters (e.g., the correct width of the grip aperture) but also to select movement-related information from the perceptual environment (e.g., the size and orientation of the cup's handle). Moreover, the context in which the grasping action is performed may also have an impact on both the action performance and the prevailing perceptual selection processes of the agent. If the cup is placed among other cups of different sizes and handle orientations, selection might be more difficult as compared to when the cup would be placed among plates. In the first case, the context varies along at least two dimensions that are *relevant* for grasping a cup (size and orientation of handles). In the second case, the cup is embedded in a homogeneous context also consisting of dimensions *irrelevant* for grasping a cup (breadth of plates). Therefore, the two environmental contexts might result in different processing speed of the environmental characteristics.

The idea that human perceptual selection is tuned to action planning has been inspired by the ideomotor perspectives of, e.g., James [1] and, most prominently, Greenwald [2]. Ideomotor theorists postulate that action planning consists in representing desired action goals in the form of anticipated sensory consequences of the planned actions. This sort of representational content is possible thanks to human life-long experience with consequences of particular actions (e.g., if one knocks on a door, they will hear a knocking sound and feel pressure on their finger joints). Thanks to acquired experience with sensory effects of actions, human perceptual system is able to veridically evaluate if the actual consequences of particular actions match the desired ones (if one knocks on the door and does not hear the desired sound, they might need to increase the force of the knocking action), for a similar approach see forward models of e.g. [3]. Such a view on human action planning implies that action systems and perceptual/sensory systems in the brain must be tightly coupled to allow for representing action in the form of its anticipated sensory consequences. Hence, already at the action planning stage, perceptual selection systems need to be tuned to what is relevant to the planned action type. This has been one of the main postulates of the selection-for-action view [4], premotor theory [5] or common coding perspectives [6, 7].

Several authors have provided empirical evidence for the ideas of close coupling between action and perception in the human brain, e.g., [6–8]. Recent neuroimaging techniques [9] and the discovery of mirror neurons, e.g., [5] also provide evidence for close coupling of perception and action. Although some authors postulate that the fast "how/where" pathway for action is separate from the slower "what" pathway for perception and conscious processing [10], recent evidence shows much closer interactions [11]. A prominent example of evidence for such a fast neuronal route linking perception with action was described by Humphreys and Riddoch [12]. The authors examined a patient MP who suffered from unilateral neglect. Following damage to the fronto-temporal-parietal regions of the right cerebral hemisphere, the patient was unable to direct his attention to objects presented in his left visual field—a symptom typical for unilateral neglect patients. When presented with a simplified visual search task, he was impaired in finding objects defined by their perceptual features when they were presented in his neglected field. For example, when asked to detect a *red* object, he was unable to detect a red cup. Interestingly, when the targets were defined by action-related features of the objects (e.g., "look for something to drink from")—so-called *affordances* [13]—the patient's performance improved markedly also in the neglected field. This indicates that there might be an efficient link between action and perception, a so-called *pragmatic processing route* that facilitates extraction of action-relevant features of the environment. Similar results supporting the idea of representation of object affordances and a pragmatic processing route were obtained in experiments involving healthy participants, e.g., [14–16].

1.1 Background of the Present Study

Recent studies demonstrated that visual selection processes are highly dependent on intended action types [17] and that the perceptual system can bias processing of action-relevant features if they are congruent with the performed action [18]. Wykowska, Schubö and Hommel [19] conducted a series of experiments in which they observed action-related biases of visual perception already at early stages of processing. Importantly, these studies showed action-perception links in a situation where action and perception were completely unrelated and decoupled but had to be performed in close temporal order. Participants' task was to detect a visually presented target (visual search task) while they were preparing for a grasping or a pointing movement. The target could be an object that differed from other objects in size or in luminance. Wykowska and colleagues found that performance in the visual search task was influenced by the intended movement although the movement was only planned at the time of completion of visual search task. That is, detection of perceptual dimension (size or luminance) was better when accompanied by the preparation of a congruent movement (e.g., grasping for size and pointing for luminance) as compared to the preparation of an incongruent movement. These results indicate a close link between action and perception that merely coincide in time. Moreover, as the perceptual task targeted at early stages of perceptual processing, the authors concluded that action preparation affected perceptual processing at the level of early mental representations.

The aim of the present study was to investigate not only the impact of action planning on perceptual processes but also the impact of the context in which the action occurs. The action context might influence not only behavior and performance of an agent [20] but also the agent's perceptual stages of processing. So far, there has been a growing body of evidence stressing the impact of perceptual context on perceptual and cognitive processing, e.g., [21–27]. However, to our knowledge, not much work has been done on how action contexts influence perception. Our goal was to fill the gap in this particular type of a research question.

2 Experimental Paradigm

We designed a paradigm in which participants were asked to perform a perceptual task (an efficient visual search task) in close temporal order to a motor task. The visual search task consisted in detecting the presence of either a size or a luminance target (in separate blocks of trials) embedded in a set of other, distracting items. The motor task consisted in grasping or pointing to a specific object on the Movement Execution Device (MED). The motor task should be planned before the visual search task but executed only after the search task was completed. In other words, while participants were performing the visual search task, the grasping or pointing movement was prepared but not yet executed. The action context consisted of objects of various sizes and surface luminance that were mounted on the MED, see Fig. 2. The action context could either be "simple" in which the objects varied along only one dimension (luminance) or "complex" in which the objects varied along two different dimensions: size and luminance. We reasoned that if action context has an impact on perceptual processing then behavioral performance of participants should be better (e.g., shorter reaction times) in the simple context condition as compared to the complex context.

2.1 Participants

Nineteen paid volunteers (16 women, mean age: 23.3) took part. Two participants were left-handed, all had normal or corrected to normal vision. All participants were experienced with other visual search experiments but they were not informed about purposes of this particular experiment. The experiment was conducted with the understanding and consent of each participant. Participants were asked to perform a practice session on the day preceding the experiment. The practice session consisted of the motor task only, without the visual search task.

2.2 Stimuli and Apparatus

The search display contained 28 grey circles of 1.7° of visual angle; 15 cd/m^2 of luminance positioned on three circular arrays with a diameter of 6.8°, 4.8°, and 2.8°. The target could appear on one of four positions on the middle circular array at the upper left/right or lower left/right from the middle point. The target was defined either by luminance (lighter grey: 58 cd/m^2, cf. Fig. 1, left) or by size (smaller circle, 1.1° cm of visual angle, cf. Fig. 1, right).

Below the computer screen, the movement execution device (MED) was positioned. It consisted of a 43 cm × 54 cm × 13 cm box containing eight holes positioned also on a circular array of 22.2° of visual angle. Round plastic items that could vary in luminance and size each covering a LED could be attached and detached from the box (cf. Fig. 2). For the purpose of this experiment, we used the following combination of objects:

(1) The "simple" action context (luminance dimension only) consisted of four grey (0.6 cd/m^2), medium-sized (3.7° of visual angle) objects and four objects that differed in luminance: two being darker (0.17 cd/m^2) and two being lighter (2.1 cd/m^2) than the standard elements (cf. Fig. 2, left).
(2) The "complex" action context (luminance and size dimensions) consisted of four grey (0.6 cd/m^2), medium-sized (3.7° of visual angle) items and four grey items

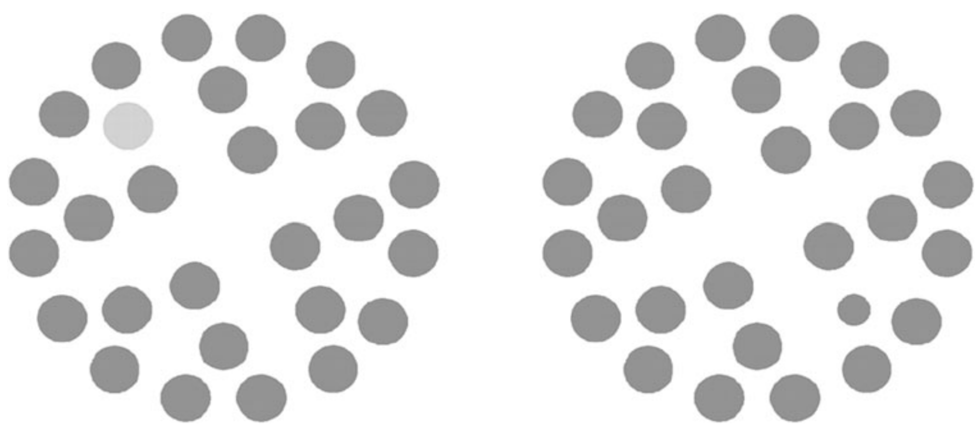

Fig. 1 Examples of visual search displays that contained a luminance target (*left*) or a size target (*right*)

Fig. 2 Movement execution device (MED). Simple action context (*left*) consisting in objects that vary along the luminance dimension (LED lighting up and the objects on MED are all defined by luminance) and complex action context (*right*) consisting in objects that vary along the luminance dimension (LED lighting up) and size dimension (items on MED)

that differed in size, two being smaller (2°) and two being larger (5.4°) than the standard elements (cf. Fig. 2, right). When the MED elements were lit up, their luminance values were equal to: 5 cd/m^2 for the darkest item, 45 cd/m^2 for the medium-light item and 99 cd/m^2 for the lightest item. The luminance values were measured with Minolta Chroma meter CS-100. MED was connected to the experimental computer via an LPT port and was controlled by the computer receiving signals at which moment which particular LED (out of the 8 attached LEDs) should light up and for how long it should remain lit.

2.3 Procedure

Participants were seated at 80 cm distance from the computer screen on which the visual search task was presented. The MED was positioned below the computer screen, in front of the participants (cf. Fig. 3). Each experimental trial began with a fixation cross displayed for 500 ms. Subsequently a movement cue (cf. Fig. 4) appeared for 1000 ms. Participants were instructed to prepare for the movement but not execute it until a signal from the MED would appear. The cue was followed by a blank display (500 ms) and, subsequently, by the search display presented for 100 ms. A blank screen followed the search display and remained on the computer screen until participants responded to the search task. Upon response to the search task, one of the LEDs on the MED lit up for 300 ms. This event constituted the signal for participants to execute the prepared movement, i.e., to either point to or grasp the object that lit up. After the experimenter registered the participants' movement with a mouse key, a new trial began (cf. Fig. 4).

Fig. 3 Experimental setup. Participants were asked to perform the movement task (grasping or pointing) on the MED with their left hand. The movement ended when participants returned with their hand to the starting position (space bar on the keyboard). The search task was to be performed with the right hand on the computer mouse (target present: left mouse key; target absent: right)

Participants were asked to be as fast and as accurate as possible in the search task. Speed was not stressed in the movement task. They were instructed in detail about their task before the experiment started. The experiment consisted of one practice block and 8 experimental blocks with 112 trials each. During the practice block, participants trained the movement task alone and subsequently, the detection task (one block for luminance and one for size) together with the movement task. The two target types (size and luminance) were always presented in separate blocks and also the configuration of objects on the MED was changed block-wise. The order of blocks was balanced across participants. The movement task was randomized within each block of trials.

2.4 Data Analysis

The analysis focused on the impact of context complexity on perceptual processing. Only correct movement trials were taken into analyses of both reaction times (RTs) as well as error rates in the search task. Prior to RT analysis, incorrect search trials, trials below 150 ms and exceeding 1200 ms, as well as other outliers in the search task (+/−3 SD from the overall mean of RT for each participant separately) were excluded from further analyses.

From the remaining data, individual mean RTs and mean error rates in the search task were submitted to analysis of variance (ANOVA) with: *target presence* (target present vs. absent), *target type* (luminance vs. size), *movement type* (point vs. grasp), and *action context* (one vs. two dimensions) as within-subject factors. If the environmental action context of a movement has an impact on perceptual processing of action-relevant characteristics, then the simpler context (varying along one dimension) should yield better performance in the search task as compared to the context consisting of two-dimensions.

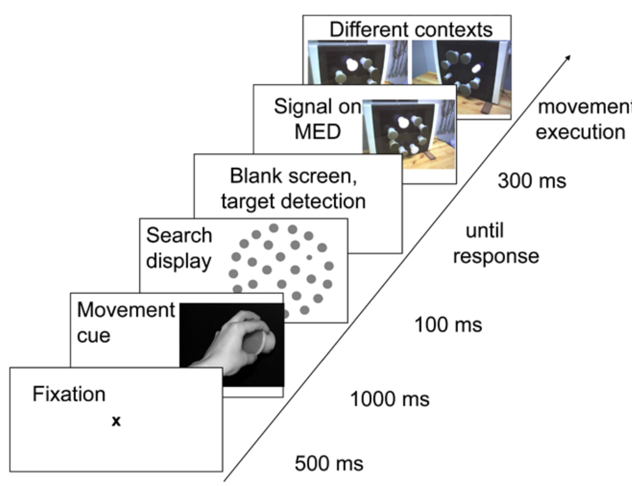

Fig. 4 A trial sequence. *Each box* represents an event of a trial (e.g., presentation of a fixation cross, of a movement cue). Presentation time for each of the event is shown *on the right side of the figure*

3 Results

The statistical analysis showed that action context had a significant effect on search performance, $F(1, 18) = 5.5$, $p < .05$ indicating that the complex action context elicited slightly longer RTs in the search task ($M = 398$ ms, $SEM = 15$) as compared to the simple context ($M = 392$ ms, $SEM = 16$). Moreover, this effect was more pronounced for the grasping movement (simple context: $M = 390$ ms, $SEM = 15$; complex context: $M = 398$ ms, $SEM = 15$) as compared to the pointing movement (simple context: $M = 395$ ms, $SEM = 15$; complex context: $M = 398$ ms, $SEM = 16$), as indicated by the significant interaction of movement type and action context, $F(1, 18) = 5$, $p < .05$ (cf. Fig. 5).

Analogous analysis for error rates in the search task showed no significant effect, all $p > .5$. However, the small absolute differences were in line with the effects in RTs (cf. Table 1).

Moreover, the present results replicated previous findings [16] by observing the influence of intended movements on perceptual dimension: interaction of movement type and target type was significant, $F(1, 18) = 5$, $p < .05$ indicating again improvement in visual search performance for congruent movements as compared to incongruent movements: Size detection was faster when a grasping movement was prepared relative to the size-incongruent movement (pointing).

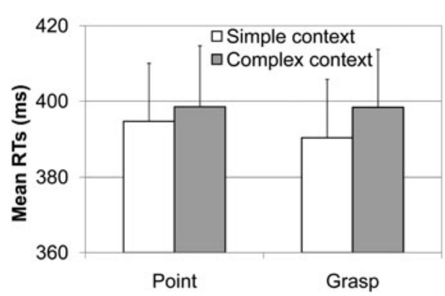

Fig. 5 Mean reaction times in the search task as a function of movement type and action context. Error bars represent standard errors of the mean (SEM)

Table 1 Mean error rates in the search task as a function of movement type and action context

Action context	Mean error rates (%) and SEM (in brackets) Movement type	
	Point	Grasp
Simple context	2.85 (0.5)	2.92 (0.7)
Complex context	2.86 (0.6)	2.96 (0.8)

4 Discussion

The aim of the present study was to investigate the influence of action context on perceptual processes during movement preparation. Results showed that action context influences already early stages of processing, namely visual selection. We believe that such results should be considered in designs of artificial systems for various reasons. First, current robotics systems have a limited computing power and, therefore, need to have selection mechanisms implemented. When designing such selection mechanisms, it might prove beneficial to draw inspiration from natural solutions of the brain. Moreover, efficient and safe human-robot interaction requires robot behaviors that are intuitive for human users. Robot behavior, in turn, might become intuitive only when it is based on similar to human processing architecture. Finally, making use of the knowledge about the workings of the human perceptual and selection mechanisms might allow for design of interfaces that are better suited to human needs. Application to robotics of findings from experimental psychology, such as the present ones, will be further discussed in the following section of the paper.

Results of the present study showed that target detection in the visual search task was faster when the context in which the movements were executed consisted of only one dimension (luminance) as compared to two dimensions (size and luminance). Interestingly, the action context effect was more pronounced for the more complex movement (grasping) as compared to the simpler pointing movement. Importantly, the design of the present experiment made the movement execution (together with its context) completely decoupled from the perceptual task. Therefore, the influences of action context on performance cannot be attributed to late stages of the whole processing stream, i.e., as perceptual consequences of the movement execution. Rather, the present results show that action context influences already early perceptual processing. Such influence suggests that the perceptual system makes use of environmental hints when selecting action-relevant features of the environment. It is important to note that the present results were obtained in a rather restricted laboratory setting where all parameters were strictly controlled so that other factors such as prior knowledge or experience with the respective action context could have no influence on the results. This might have resulted in relatively small absolute differences in reaction times. These effects would probably become larger once they were examined in more natural settings, such as, for example, a real kitchen scenario.

The present findings can be interpreted in a broad, evolutionary context. Humans must have developed mechanisms that optimize their interaction with the environment. As described above, the means for optimizing such interaction might be to represent planned actions in the form of their sensory consequences: if the actual consequences match the desired, action has been successful. If not, certain movement parameters need to be adjusted. This implies that the perceptual system needs to be tuned to the action-relevant environmental characteristics. A common representational code for action and perception [6] might have turned out to be a convenient and parsimonious way of allowing fast and efficient action-perception transition. Such common code implies bidirectional links that can be used in such a way that action planning can influence perception no less than perception has an impact on action preparation. Human perceptual system works in a selective manner since the sensory systems receive abundance of data that are not all necessarily relevant, e.g., [28–30]. To filter out the irrelevant information, human control mechanisms might enhance processing of the relevant characteristics and suppress those neural responses that are related to irrelevant characteristics. For example, if one searches for a person dressed in a red shirt within a crowd of people, they will prioritize processing of the red colour and suppress processing of green or blue that are not relevant to the task of finding a person dressed in red. Similarly, when preparing for a certain movement type like, for example, grasping a cup, humans might prioritize processing of grasping-relevant characteristics (such as the diameter of the cup) and attenuate the neural responses related to, for example, colour. Such biased processing might be achieved with the means of a mechanism that weights perceptual dimensions with respect to a desired goal. Hommel et al. [6] have termed such action-related weighting mechanism “intentional weighting”.

The present results suggest that the interactions between perception and action might be even more complex. Our results showed that not only intended actions play a role in perceptual selection processes but also environmental contexts in which the actions are performed. If the context varies along several dimensions, perceptual selection is presumably more difficult. When humans are preparing a movement, they try to optimize their performance through efficient selection of perceptual information that might be relevant for that movement. This optimization process might be especially important for complex movements as they are more difficult. Therefore, humans might benefit from the environmental setting (i.e., a simple action context) more in case of more difficult interaction (more complex movement) as compared to a movement that is simple enough not require such a level of optimization.

5 Implication for Robotics

As argued above, humans have developed certain mechanisms that allow for optimization of perception mechanisms given the task at hand. These principles can be applied quite directly to the design of robotics systems. The observed behavior on people support the design of modular perception systems, that can run only the algorithms relevant to the current action and using only the necessary sensory streams. Limited computing power is not the only reason for such a design. Horswill [31] proposed that by exploiting the constraints of the tasks and the environment, a general perception problem can be substituted by a more specific one, and it is possible to prove that the specific method works more robustly and efficiently than the general one as long as the constraints hold. On a similar line, Kei Okada and colleagues described action-specific behavior verification routines: for example, an algorithm to detect if the water is running after opening a faucet, or to observe the level of liquid in a glass when transporting it [32].

The basic idea is to choose the right variables to observe depending on the action. If the scene is very complex, with hundreds of possible sources of information (i.e. many objects and people around the robot, images, sonar and laser data, torque, velocity, and angle information from the arms, sensory stream from the robot fingers), then it becomes essential to carefully chose which ones to observe for a specific task. Consider the ideas of Itti and Arbib [33] where they propose to focus the attention on the minimal subscene, that only includes the relevant objects, so that the robot should assign more resources to tracking the object that is being manipulated and the nearby obstacles.

In our research we apply the ideas of action- and context-guided perception to controlling autonomous robots performing everyday manipulation tasks [34, 35]. Figure 6 shows our mobile manipulation platform performing pick-and-place tasks in a kitchen environment. The robot is equipped with sensors including a stereo camera setup, a 3D laser scanner, a time-of-flight camera, two laser scanners for navigation, and a micro-camera-based fingertip sensor. Each one of the mentioned components delivers its own sensory stream, with a bandwidth ranging from approx. 100 KiB/s to 1000 MiB/s. The perception system [36] performs a variety of perceptual tasks including perception of objects and interpretation of scenes. In order to perceive objects the system can detect, find, classify, localize, recognize, (geometrically) reconstruct models in its sensor view. It also employs multiple methods for performing similar perceptual tasks. Thus, the system has algorithms for the tasks that work on CAD models, point cloud information, SIFT, or feature classification, which enables it to apply the methods best suited for the particular situation it is confronted with.

Action- and context-guided perception in our control system is primarily applied in two ways. First, the robot assigns computational resources in perception to where they are needed to complete the robot tasks successfully. Second, the system uses the knowledge about the task and action context in order to specialize the perception tasks. When the

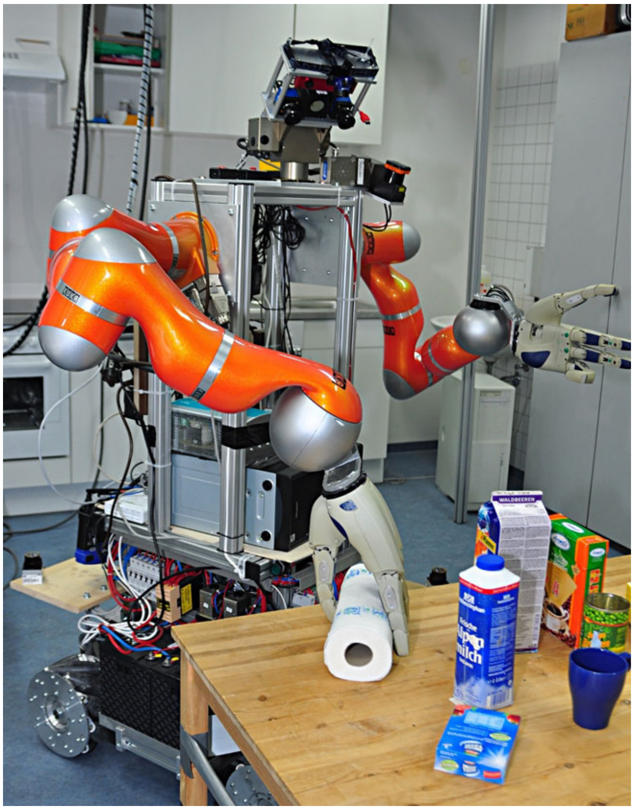

Fig. 6 Our robot for manipulation in human environments: TUM-Rosie, equipped with two KUKA LWR-4 arms and DLR-HIT 4-finger hands. It performs pick and place tasks in a kitchen at the Intelligent Autonomous Systems Laboratory, TUM. The high amount of sensory data and limited computing power on board of the robots make it necessary to carefully choose the specific perception algorithms as needed for the current action

high-level plan of the robot decides to execute a particular task, for example to pick up a tea box, then the appropriate perception algorithms are chosen. In this case, since we have a detailed CAD model of the object we want to manipulate, then the CAD based algorithm [37] is used. When the robot needs to grasp unknown objects, then an approach based on point clouds [38] is chosen.

Different perception systems are used during the pick and place procedure (cf. Fig. 7). After the objects are detected, the processing power is devoted to the sensory information coming from the robotic arm and hand to correctly grasp the object. For example, after choosing a desired grasp for the DLR-HIT robotic hand, the grasping is supervised using a supervised-learning algorithm on torque and angle information from the robot fingers. A set of grasping controllers was developed, one for each type of basic grasp movement (e.g. power grasp, precision pinch, and others). After learning a sensory model from several examples, the system can quickly differentiate between good and bad grasps. It works with a high accuracy without being CPU intensive because it processes a sensory stream with low bandwidth (only what

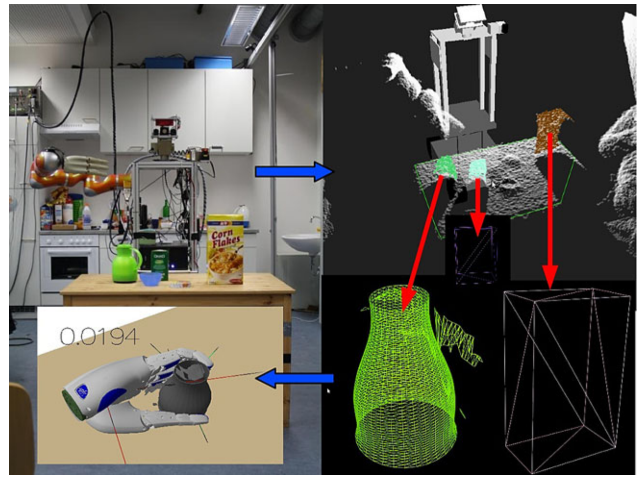

Fig. 7 Different perception mechanisms are activated depending on the task at hand. (*Left-up*) Object detection based on CAD models. (*Right*) Detection of unknown objects for grasping using time of flight data. (*Left-bottom*) Object reconstruction using symmetric characteristics

is task relevant). When executing a grasp, only the associated model is fed with the sensory stream coming from the hand, and the system can monitor the success of the action. This design also complies with the ideas of Wolpert and Kawato of multiple interconnected forward and inverse models [3]. On the other hand, a program that uses camera images to detect the object and the hand would be many times more complex and not perform better in terms of accuracy or speed. The same idea applies to attention: We keep a representation of the objects as simple as possible (point clouds or texture) and only do a more detailed 3D object reconstruction when we need it to grasp the objects using a geometrical grasp planner.

To give an example: the robot wants to set the table, but if there is a human in the scene, then it should also track the movements and actions of the human. If the person is around, it has to spend computation resources on perceiving what the person does. If not, these mechanisms are simply deactivated. For a more comprehensive discussion of the algorithms mentioned, please look at [34–38]. Our methods validate the results of this paper. In our ongoing research, we further investigate this line of research.

Another line of possible application to robotics of empirical results obtained in the area of cognitive psychology, such as the present ones, is interface design. Recent work has suggested that an interface making use of the close action-perception link in the human processing system might have a beneficial effect on human cognitive capacities. Cabibihan and colleagues [39] developed a mobile platform for online communication. Participants communicated with an experimenter via Skype (Skype Technologies, Luxembourg) either in a standard static mode through a computer screen or in a dynamic mode with the use of robot arms that could point to a given direction. An experimenter described a spatial layout of items in imaginary rooms either using a verbal description (e.g., to the left/right; behind)—static mode, or using the robot arms pointing to a direction and using such a vocabulary as "there"—dynamic mode. The task of the participants was to subsequently recall the positions of the items. Performance was better in the dynamic condition as compared to the static mode. Although the authors interpret the results as showing better performance in a condition with more rich social cues, we claim that these results might as well show that the dynamic mode of the platform uses the "direct pragmatic route" [12] that links action and perception systems in the human brain and as such, facilitates cognition. In any case, this piece of research is an excellent example how design of interfaces can profit from results in experimental psychology.

Acknowledgements This study was supported by the Deutsche Forschungsgemeinschaft (SCHU 1330/5-1 and Project #301, Excellence Cluster 'Cognition for Technical Systems', CoTeSys).

References

1. James W (1890) Principles of psychology. Holt, New York
2. Greenwald AG (1970) Sensory feedback mechanisms in performance control: with special reference to the ideo-motor mechanism. Psychol Rev 77:73–99
3. Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. Neural Netw 11:1317–1329
4. Allport A (1987) Selection for action: some behavioral and neurophysiological considerations of attention and action. In: Heuer H, Sanders AF (eds) Perspectives on perception and action. Erlbaum, Hillsdale, pp 395–419
5. Rizzolatti G, Craighero L (2004) The mirror-neuron system. Ann Rev Neurosci 27:169–192
6. Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The theory of event coding (TEC): a framework for perception and action planning. Behav Brain Sci 24:849–937
7. Prinz W (1997) Perception and action planning. Eur J Cogn Psychol 9:129–154
8. Craighero L, Fadiga L, Rizzolatti G, Umiltà CA (1999) Action for perception: a motor-visual attentional effect. J Exp Psychol Hum Percept Perform 25:1673–1692
9. Schubotz RI, von Cramon DY (2002) Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. Neuroimage 15:787–796
10. Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15:20–25
11. Rossetti Y, Pisella L, Vighetto A (2003) Optic ataxia revisited: visually guided action versus immediate visuomotor control. Exp Brain Res 153:171–179
12. Humphreys GW, Riddoch MJ (2001) Detection by action: neuropsychological evidence for action-defined templates in search. Nature Neurosci 4:84–89
13. Gibson EJ (1977) The theory of affordances. In: Shaw RE, Bransford J (eds) Perceiving, acting and knowing. Erlbaum, Hillsdale, pp 127–143
14. Tucker R, Ellis M (2001) The potentiation of grasp types during visual object categorization. Vis Cogn 8:769–800
15. Grèzes J, Decety J (2002) Does visual perception of object afford action? Evidence from a neuroimaging study. Neuropsychologia 40:212–222
16. Grafton ST, Fadiga L, Arbib MA, Rizzolatti G (1997) Premotor cortex activation during observation and naming of familiar tools. NeuroImage 6:231–236
17. Bekkering H, Neggers SFW (2002) Visual search is modulated by action intentions. Psychol Sci 13:370–374
18. Fagioli S, Hommel B, Schubotz RI (2007) Intentional control of attention: action planning primes action related stimulus dimensions. Psychol Res 71:22–29
19. Wykowska A, Schubö A, Hommel B (2009) How you move is what you see: action planning biases selection in visual search. J Exp Psychol Hum Percept Perform 35:1755–1769
20. Turner RM (1998) Context-mediated behavior for intelligent agents. Int J Hum-Comput Stud 48:307–330
21. Chun MM, Jiang Y (1998) Contextual cueing: implicit learning and memory of visual context guides spatial attention. Cogn Psychol 36:28–71
22. Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. Psychol Rev 96:433–458
23. Kunar MA, Flusberg S, Horowitz TS, Wolfe JM (2007) Does contextual cuing guide the deployment of attention? J Exp Psychol Hum Percept Perform 33:816–828
24. Olson IR, Chun MM, Allison T (2001) Contextual guidance of attention. Human intracranial event-related potential evidence for feedback modulation in anatomically early, temporally late stages of visual processing. Brain 124:1417–1425
25. Schankin A, Schubö A (2009) Cognitive processes facilitated by contextual cueing. Evidence from event-related brain potentials. Psychophysiology 46:668–679
26. Schubö A, Schröger E, Meinecke C (2004) Texture segmentation and visual search for pop-out targets: an ERP study. Brain Res Cogn Brain Res 21:317–334
27. Schubö A, Wykowska A, Müller HJ (2007) Detecting pop-out targets in contexts of varying homogeneity: investigating homogeneity coding with event-related brain potentials (ERPs). Brain Res 1138:136–147
28. Bundesen C (1990) A theory of visual attention. Psychol Rev 97:523–547
29. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Ann Rev Neurosci 18:193–222
30. Wolfe JM (1994) Guided search 2.0: a revised model of visual search. Psychon Bull Rev 1:202–238
31. Horswill I (1994) Specialization of perceptual processes. PhD thesis, Massachusetts Institute of Technology, Cambridge
32. Okada K, Kojima M, Sagawa Y, Ichino T, Sato K, Inaba M (2006) Vision based behavior verification system of humanoid robot for daily environment tasks. In: 6th IEEE-RAS international conference on humanoid robots, Humanoids 2006, pp 7–12
33. Itti L, Arbib M (2006) Attention and the minimal subscene. In: Arbib MA (ed) Action to language via the mirror neuron system. Cambridge University Press, Cambridge, pp 289–346
34. Beetz M, Stulp F, Esden-Tempski P, Fedrizzi A, Klank U, Kresse I, Maldonado A, Ruiz-Ugalde F (2010) Generality and legibility in mobile manipulation. Auton Robots J (special issue on mobile manipulation)
35. Beetz M, Jain D, Mösenlechner L, Tenorth M (2010) Towards performing everyday manipulation activities. Robot Auton Syst
36. Beetz M, Blodow N, Klank U, Marton Z, Pangercic D, Rusu R (2009) COP-MAN perception for mobile pick-and-place in human living environments. Workshop, IROS
37. Klank U, Pangercic D, Rusu R, Beetz M (2009) Real-time CAD model matching for mobile manipulation and grasping. In: 9th IEEE-RAS international conference on humanoid robots, Paris, France, December 7–10

38. Maldonado A, Klank U, Beetz M (2010) Robotic grasping of unmodeled objects using time-of-flight range data and finger torque information. In: IEEE/RSJ international conference on intelligent robots and systems, IROS, 2010 (accepted for publication)
39. Cabibihan J-J, So WC, Nazar M, Ge SS (2009) Pointing gestures for a robot mediated communication interface. In: Xie M et al (eds) ICIRA 2009. LNAI, vol 5928. Springer, Berlin, pp 67–77

Agnieszka Wykowska received her M.Sc. in Neuro-Cognitive Psychology from Ludwig Maximilian University (LMU) in Munich and a M.A. in Philosophy from the Jagiellonian University, Krakow, Poland. She completed her Ph.D. in 2008 with honors in the Department of Psychology, LMU. Since then, she has been employed as a post-doc in the Department of General and Experimental Psychology, LMU, being involved in the German cluster of excellence COTESYS (Cognition for Technical Systems) as well as in the DFG Research Grant titled "How top-down mechanisms, action intentions and bottom-up processes guide visual attention". Her research interests are human visual attention, interactions between action and perception, as well as human-robot interaction. In her research, she uses EEG/ERP methodology combined with behavioural measures.

Alexis Maldonado Herrera received his Master's (Licenciatura) degree in electrical engineering with emphasis in digital control from the Universidad de Costa Rica, San José, Costa Rica in 2004. After working as a lecturer at the same university and in automation projects for the local industry, he became a DAAD (German Academic Exchange Service) scholar and joined the laboratory of Prof. Michael Beetz at the Technische Universität München, where he currently works towards his Ph.D. degree. His interests include robotic manipulation in unstructured environments, system integration and contributing to free software and hardware robot projects.

Michael Beetz is a professor for Computer Science at the Department of Informatics of the Technische Universität München and heads the Intelligent Autonomous Systems group. He is also the vice coordinator of the German cluster of excellence COTESYS (Cognition for Technical Systems). Michael Beetz received his diploma degree in Computer Science with distinction from the University of Kaiserslautern. He received his M.Sc., M.Phil., and Ph.D. degrees from Yale University in 1993, 1994, and 1996 and his Venia Legendi from the University of Bonn in 2000. Michael Beetz was a member of the steering committee of the European network of excellence in AI planning (PLANET) and coordinating the research area "robot planning". He is associate editor of the Artificial Intelligence journal. He was also principal investigator of a number of research projects in the area of AI-based robot control. His research interests include plan-based control of autonomous robots, knowledge representation and processing for robots, integrated robot learning, and cognitive perception.

Anna Schubö is a senior lecturer at the Department of Psychology of the Ludwig-Maximilian University in Munich. She received her Diploma in Psychology from LMU and her Ph.D. from the Max-Planck Institute for Psychological Research in Munich. After a Postdoc in Erlangen, she obtained the venia legendi in Psychology in Munich. Anna is interested in various fields of Cognitive Psychology, Cognitive Neuroscience and the implementation of Cognition in Robotics. Her work concentrates on visual perception, the allocation of attention, motor control processes and on the inter-relation of human perception and action. Since 2006, she is principal investigator in the Cluster of Excellence "Cognition for Technical Systems" (CoTeSys) and works on cognitive capabilities in human and robot action.