Does Observing Artificial Robotic Systems Influence Human Perceptual Processing in the Same Way as Observing Humans?

Agnieszka Wykowska[1], Ryad Chellali[2],
Md. Mamun Al-Amin[1], and Hermann J. Müller[1],

[1] General and Experimental Psychology Unit, Dept. of PSychology,
Ludwig-Maximilians Universität, Leopoldstr. 13, 80802 Munich, Germany
[2] Instituto Italiano di Tecnologia-PAVIS, Via Morego, 30,
16165 Genova, Italy
agnieszka.wykowska@psy.lmu.de, ryad.chellali@iit.it,
bd_pharmacy@yahoo.com, hmueller@psy.lmu.de

Abstract. Humanoid robots are designed and shaped to have physical bodies resembling humans. The anthropomorphic shape is aiming at facilitating interactions between humans and robots with the ultimate goal of making robots acceptable social partners. This attempt is not very new to roboticists and there is an increasing body of research showing the importance of robots' appearance in HRI; the Uncanny Valley proposed in the 70's [1] is however still an open problem. Our aim in this contribution is to examine how human perceptual mechanisms involved in action observation are influenced by the external shape of observed robots. Our present results show that observing robotic/cartoon hands performing grasping/pointing movements elicits similar perceptual mechanisms as observing other humans. Hence, it seems that observing actions of artificial systems can induce similar perceptual effects as observing actions of humans.

Keywords: Attentional selection, Perceptual processing, Human-Robot Interaction.

1 Introduction

The field of humanoid robotics aims at designing artificial agents that will help people in their daily lives and as such, these robots *should* be part of humans social sphere [20]. This assumption is strong and its realization depends on whether or not humans will consider humanoids as socially accepted partners or simple machines. If robots are to be perceived as machines or simple automata, it is enough to consider – for robotic designs - only the functions they support and the ways to access these functions. However, For a robot to be perceived as a social partner with which natural interactions are possible, mechanisms underlying social perception in the human mind need also to be taken into account.

S.S. Ge et al. (Eds.): ICSR 2012, LNAI 7621, pp. 327–337, 2012.

Reproducing the human appearance and/or behaviors is the very common argument used by humanoid robotics researchers in designing humanoid robots. It is believed that anthropomorphic shapes should enable sensory-motor mappings and thus facilitate social interactions. However, this argument is far from finding strong justifications and measurable benefits and the quest is still an open problem.

The robot's shape is assumed to be one of the critical characteristics/limit the impressions/perception one can have about the agent. The *uncanny valley hypothesis* has been proposed in the 70's [1] and addressed from different angles. To date, many contradicting results have been found, preventing from drawing clear and useful conclusions about the Mori's intuition and more practically, no design guidelines have been derived.

For example [21] has suggested an "uncanny cliff" instead of an "uncanny valley", and Mac Dorman and colleagues [22] reported data suggesting that there are many factors which can influence perception of a robot as strange or eerie, apart from the *human-likeness* dimension (see also [23] for a similar account).

In our work, the core idea we used relies on the well-established paradigm about action-perception coupling in humans when observing other humans acting. We aimed at investigating two issues: 1) the effects of observing artificial agents on human perceptual system; 2) determining whether the shape of the used artificial agent (a humanoid robot or a cartoon arm) influences the human perceptual system.

Action and Perception Domains – Are They Distinct?

Traditional approaches viewed human cognitive system processing information in a unidirectional, stage-like manner [2]. It has been postulated that humans first process perceptual information, and then selected information is transmitted to memory and action planning sub-systems, to be eventually used for action execution. More recent theoretical frameworks, however, have stressed the idea that action and perception are more directly coupled to the extent that they share a common representational code [3,4]. If action and perception are so tightly coupled, then such coupling should allow bidirectional mutual influences. This would imply that information processing stream does not flow from perception to action planning and execution, but action planning influences already the earliest (perceptual) processes to the same extent as perception influences action planning [5]. In [6], participants were asked to prepare a grasping or pointing movement while simultaneously performing a visual search task for action-congruent or incongruent target dimensions. Performance was found to be better for the congruent action-perceptual target pairs, relative to the incongruent pairs. The observed *congruency effects* were due to the way the brain represented planned actions and perceived stimuli, and not due to any perceptual or motoric overlap between the two tasks [7].

Neuronal Mechanisms of the Action-Perception Links

The so-far described empirical evidence was based on behavioural studies with human participants. The idea of a close coupling between action and perception has been supported also with a growing body of evidence collected with the use of neuroimaging techniques. For example, Schubotz and von Crammon [8] carried out a study in which participants observed sequences of differently sized disks or listened to sounds of various pitches. The results showed activation of premotor areas usually involved in hand movements when sequences of visually presented disks were observed; analogously, premotor areas involved in articulation were activated when participants judged the auditory stimuli.

More specifically, the so-called *mirror neurons* [9] are perhaps the most straightforward evidence for tight coupling between perception and action. Indeed, these neurons are active not only when one is planning a particular movement type but also when one only observes this type of movement being performed by others. Mirror neurons have been claimed to have the function of action understanding and imitation learning [10]. The complex functionality of the mirror neuron system might also include the foundations for social [11] and emotional [12] behaviors. Then the key question arises: do mirror neurons are also activated when actions of artificial systems are observed?

Aim of the Present Study

In the context of social robotics enterprise, it seems intuitive that the more human-like appearance a robot has (e.g., android robots), the closer and easier the human-robot interaction should be. Therefore, it is crucial to examine the human cognitive mechanisms underlying social interactions with human-like robots. One of such mechanisms is the action-perception coupling during action observation. In other words, the important question to be addressed is: does observing a humanoid robot performing a certain action evoke in humans similar perceptual mechanisms that would be evoked if the human observers watched other members of their species performing that action?

In the present study we addressed this question using an experimental paradigm developed by Wykowska and colleagues [6], which elicits the so-called action-perception *congruency effects*. In this paradigm, participants are typically asked to perform a perceptual task (detect a target item defined by a particular feature among other items that differ from the target with respect to that feature – a so-called visual search task) and a movement task (grasping or pointing to an object) that is signaled by a picture depicting a grasping or a pointing arm. Importantly, the target (together with the other items of the visual search task) is presented on a computer screen, while the movement is to be performed on objects that are placed below the computer screen. Moreover, the movement task is supposed to be prepared before the visual search task, but executed only after the visual search task is completed. With this paradigm typically two congruency pairs are created: 1) the target defining feature dimension of size with grasping; and 2) the target-defining dimension of luminance with pointing. It is assumed that size is a grasping-congruent feature dimension (one needs to specify the size of the to-be grasped objects, in order to adjust grip aperture) and luminance is a pointing-congruent feature dimension (luminance targets enable efficient localization of an object with a pointing-movement response). Therefore, if detection of size-defined targets in a visual search task is facilitated for the grasping condition relative to pointing; and luminance targets are detected better in the pointing condition relative to grasping, then one can conclude that congruency effects are observed.

For the present paradigm, we designed an experiment in which the pictures of human hands grasping or pointing [6,7] are substituted by cartoon hands and robot hands. If robots are capable of inducing action-perception congruency effects, those effects should be observed independent of the type of cue signaling the movement.

2 Methods

2.1 Participants

Twenty healthy volunteers (7 women; age range: 18-30 years; mean age = 24.25) took part in the experiment. Two participants were left-handed but they were standardly using the computer mouse in the same way as right-handers. The participants were naïve with respect to the purposes of this experiment. All participants had normal or corrected to normal vision and have provided informed consent regarding participation in the experiment.

2.2 The Paradigm

Participants were required to perform two types of visual search tasks (search for luminance or size targets) and two types of movement tasks (grasping or pointing an indicated cup placed below the computer screen). Target present trials are shown in the Figure 1. In luminance detection task, participants searched for one lighter circle among other circles (Figure 1A). In size detection task, participants searched for one large size circle among other same size circles (Figure 1B).

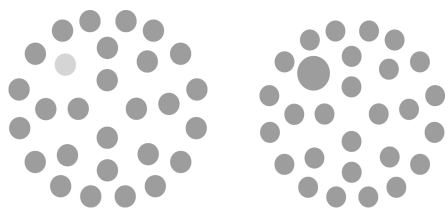

Figure 1 A Figure 1 B

Fig. 1. The display of the visual search task with luminance target (1A) and size target (1B)

In the movement task, participants first observed a cue signaling the type of movement (grasping or pointing). The cue consisted in either a cartoon of a human hand (Figure 2A, B) or a robot hand (Figure 3A, B). Only after participants completed the visual search task, were they allowed to execute the prepared grasping or pointing movement. With such a paradigm, we created two action-perception congruent pairs: i) size targets with grasping movement; and ii) luminance targets with pointing movement (according to previous studies [6,7], size is relevant for grasping and luminance is coupled with pointing). *Congruency effects* would therefore consist in faster reaction times (RTs) in the visual search task for size targets when the grasping movement is prepared, relative to the pointing movement; and faster RTs for luminance targets when the pointing movement is prepared, as compared to grasping.

2.3 Stimuli and Apparatus

Stimuli were presented on 19’ CRT screen, with a 100 Hz refresh rate placed at a distance of 100 cm from an observer. Responses were registered with a Logitech optical mouse. The whole experiment was programmed in E-Prime® (Psychology Software Tools, Inc.) run on an Intel® Core™ 2 CPU 6700 @ 2.66 Ghz, 2CPUs).

Visual Search: The visual search displays consisted in three imaginary circular arrays of 6.8°, 4.8°and 2.8° diameter, with 16, 8 and 4 visual search items, respectively. The target item was always presented in the middle circle of the array. *Size targets:* Size-target search display comprised of 28 grey circular items (1.1° of visual angle; 15 cd/m^2 of luminance). The target item was always a larger circle (1.4° of diameter). All displays were shown on a light-gray background (Figure 1). *Luminance targets:* All search items were of the same size (1.1°). The target item was always lighter (luminance: 58 cd/m^2) than the other circles. All displays were shown on a light-gray background (Figure 1B). In the target-absent trials, all items were identical (1.1° of visual angle; 15 cd/m^2 of luminance). There were 50% of target present and 50% of target absent trials.

Figure 2 A Figure 2 B

Fig. 2. Cartoon hand signaling a pointing movement (2A) and a grasping movement (2B)

Figure 3 A Figure 3 B

Fig. 3. Robotic hand signaling a pointing movement (3A) and a grasping movement (3B)

Movement Task Apparatus: The movement cues (Figure 2&3) were presented in the middle of the computer screen. The movements were to be executed on three different cups, which were made up of hard papers. They differed in size and luminance: a small white, 5 cm in diameter in the middle point; a middle grey cup, 6.5 cm in diameter in the middle point; and a large dark grey cup, 8 cm in diameter in the middle point. They were all equal in height and weight. All the cups placed minimum 25 cm away from the display monitor on the table.

The go-signal for the movement execution consisted in a yellow asterisk of 0.6° in diameter. It was presented 4.5°, 11.3°, or 17.7° from the left border of the screen signalling the to-be-grasped/pointed to paper cup situated beneath the computer screen, each cup being situated below one of the asterisk positions. The cup positions were changed after each block.

2.4 Procedure

All participants were seated in a quiet and dimly lit room with response mouse positioned under their right hand and placed on the lap (see Figure 4).

Fig. 4. Experimental setup inside the chamber

At the beginning of each trial a 300 ms fixation cross was displayed ("x" in the center of the screen). Subsequently, the movement cue was presented for 800 ms. Next, subsequent to another fixation cross (200 ms) a visual search display was presented for 100 ms (see Figure 5). The visual search display was followed by a blank screen during which participants were supposed to respond to the visual search display (target present vs. target absent). Upon the visual search response, and another fixation cross (400 ms), the go-signal for the movement execution was presented. The go-signal consisted in a yellow asterisk presented for 300 ms. Participants were to execute the prepared movement in response to the asterisk. Movements were registered by the experimenter with the use of a web camera (Microsoft LifeCam VX – 800) and a computer mouse. The next trial began subsequent to the experimenter's registration of the movement performed by the participant. Participants were instructed to respond as fast and correctly as possible in the search task. In the movement task only correctness was stressed. Participants were provided with feedback concerning their performance after each block. Visual search tasks (luminance vs. size) were blocked and the order of blocks was counterbalanced across participants. The movement types (grasping vs. pointing) were randomized within blocks. The cue types (robot vs. cartoon) were presented in two separate experimental sessions (8 blocks, 60 trials each) with the order counterbalanced across participants. Hence, each participant took part in three sessions on three separate days. The first session consisted in practicing only the movement task (15-30 min.; 5 blocks), so that the subsequent experimental sessions involving two tasks would be easier to perform.

2.5 Data Analysis

Error rates were computed for each participant in both the search task and the movement task. Prior to reaction time (RT) analysis in the search task, errors in any of the two tasks as well as outliers in the search task were excluded (+/- 3 SD from mean RT for each participant, each cue type and each task type separately). Error rate analyses in the search task were conducted on correct movement trials. Data of three subjects whose error rates in any of the search tasks were above 15% were excluded from the analyses (mean error rates for the other participants: 7%). From the remaining data, mean RTs and mean error rates were calculated and subject to an analysis of variance (ANOVA) with the within-subject factors of cue type (robot vs. cartoon) × task type (size vs. luminance) × movement type (grasping vs. pointing) × display type (target vs. blank) and a between-subject factor of cue type order (robot cues first vs. cartoon cues first).

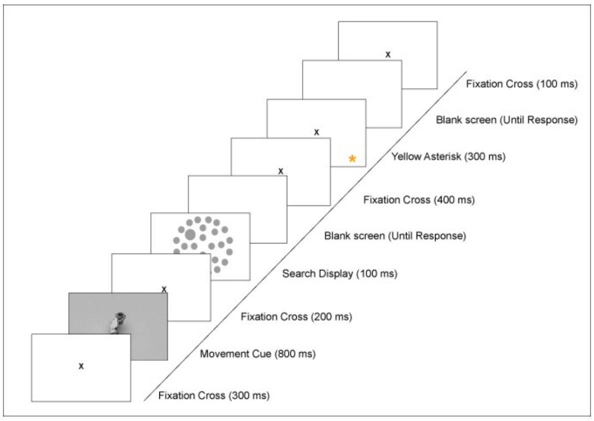

Fig. 5. A trial sequence of the present experiment

3 Results

3.1 Reaction Times

The analysis performed on the mean RT data revealed a main effect of display type, $F(1,15)=6.7$, $p<0.05$, $\eta_p^2=.3$ showing faster RTs to target trials (M = 484 ms) as compared to blank trials (M = 522 ms). Most importantly, the interaction of movement type and task type was significant, $F(1,15)=7.9$, $p=0.05$, $\eta_p^2 = .35$ (see Figure 6) and it did not depend on cue type, $p > .8$ (see Figure 7), cue type order, $p > .065$, or the combination of the two, $p > .6$. This interaction was further tested with planned comparisons, which revealed a congruency effect for luminance targets: RTs in luminance detection task were faster when participants concurrently prepared for pointing (M= 513 ms) as compared to grasping (M=520 ms), t (16) =2, $p < .05$, one-tailed. In the size detection task, the RTs showed also a tendency for the congruency effect: slightly faster RTs for grasping (M= 489 ms) as compared to pointing (M=491 ms), but this effect obviously did not become statistically significant, $p > .2$.

3.2 Error Rates

Analogous analysis on mean error rates showed only the main effect of display type, $F(1,15)=5.6$, $p<0.05$, $\eta_p^2=.2.7$ with less errors for blank trials (M= 4.3%) than for target trials (M= 7.6%), main effect of task type $F(1,15)=6.3$, $p<0.05$, $\eta_p^2=.3$ with less errors in the size task (M= 4.6%) as compared to the luminance task (7.3%). Moreover, the main effect of movement type reached the level of significance in the error rates, $F(1,15)=15$, $p<0.005$, $\eta_p^2=.5$, showing better performance in the pointing condition (M = 5.1%) as compared to grasping (M = 6.9%). The interaction between movement type and task type was not significant, p > .1.

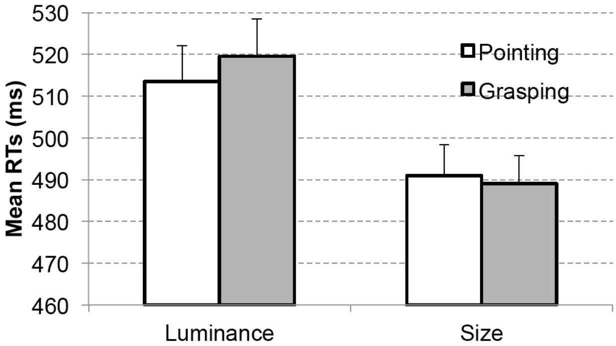

Fig. 6. Mean reaction times (RTs) as a function of task type (luminance vs. size) and movement type (pointing vs. grasping). The differences between the movement types for each of the visual search tasks are the *congruency effects*. Error bars represent within-subject confidence intervals with 95% probability criterion calculated according to the procedure described in Cousineau (2005).

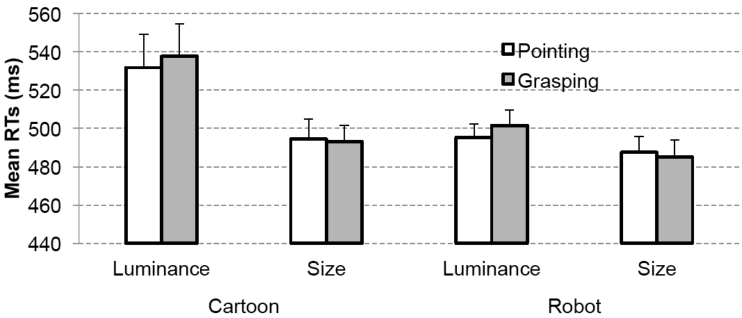

Fig. 7. Mean reaction times (RTs) as a function of movement type (point vs, grasp), task type (size vs. luminance) and two types of cues (cartoon or robot). Error bars represent within-subject confidence intervals with 95% probability criterion calculated according to the procedure described in Cousineau (2005).

4 Discussion

The aim of this experiment was to test whether observing pictures of robot hands or cartoon hands performing two types of movement (grasping or pointing) would elicit similar perceptual processes in the human observers as in the case of viewing pictures of human hands. The paradigm targeted at measuring the so-called *congruency effects* [6], which are related to better performance in a perceptual task (visual search) when a concurrently prepared movement is congruent with the dimension of the visual search target, as compared to an incongruent movement. The paradigm introduced two congruency pairs: two types of visual search targets coupled with two types of movements, i.e., one congruent pair consisted in size target + a grasping movement and the other pair consisted in luminance target + a pointing movement. Participants performed the two unrelated tasks (movement task and visual search task) concurrently.

In line with the previous studies [6,13], the present results showed congruency effects for the luminance targets and a trend for the congruency effects for size targets. Importantly for the purposes of the present experiment, the congruency effects were independent of the cue type. Hence, it seems that the basic (and perhaps implicit) perceptual effects can be elicited regardless the particular type of the observed agent. This is in line with a study by Oberman et al. [14] who found that observing both human and robot actions activated the mirror neurons in human beings. This mirror neuron activation facilitated reproduction of the same observed action. Oztop et al. [15] also found that both humans and humanoid robots elicited interference effects when being observed simultaneously with execution of movements that could be either congruent or incongruent with the observed ones. Taken together, these findings supplemented by the present results, suggest that similar perceptual processes are evoked when observing humans or humanoid robots.

It might be the case, however, that the complexity of information carried by the stimulus material might play a role in inducing congruency effects to various degrees. Hence, different degrees of congruency effects might be found when dynamic videos of actual robot arms grasping/pointing are presented, and the congruency effects are compared to a condition in which dynamic videos of human arms grasping/pointing are observed.

In addition to the effects of major theoretical interest, the present analyses revealed also that participants were faster in responding to target present trials as compared to target absent trials, which is a common finding in the visual search literature [6, 16,17] indicating different processing modes for situations when a signal is present in the visual field as compared to being absent [18, 19]. Error rates depended on the type of movement revealing better performance in the pointing condition as compared to grasping and on the type of task – luminance targets were detected with higher accuracy. This might partially explain why the congruency effects were observed only in the luminance task. It might be that congruency effects are generally observable in tasks, which are not too difficult to perform. As the congruency effects are postulated to be perceptual in nature [6], any response strategies might obscure them. Hence, in case of more difficult tasks, when response strategies come into play, the behavioral measures might not be sensitive enough to pinpoint the perceptual effects of interest.

5 Implications for Social Robots Design

Engineers design robots and sometimes they ask art designers to create agents with a nice and an appealing look to be accepted by users. As argued above, the robot is an ambiguous concept, neither a machine nor a biological being, but having an apparent intelligence that made humans consider robots as potential companions. In this study, a first step toward understanding the effects of robots morphology on human perception in HRI contexts, we aimed at objectively measuring the effects of robots morphology on the human mind. We reproduced an experimental schema that has been tested for humans observing humans: we extended it to humans observing cartoon-like hands and an anthropomorphic robotics hand. Some interesting trends have been elicited showing a “shape” effect, mainly for the *luminance* task. To our knowledge, this approach is the first to be tested in order to develop an effective metrics to measure the relationship between robots morphology and its effects on humans in terms of action-perception congruency (or sensory-motor links). This study is in its early stages and will be extended to other types of arms/hands and more dynamic stimuli to better understand the potential mapping between the robot’s morphology and the ways it affects the human companion. Two main directions will be considered: 1- movement, 2- morphology. We have already some arms prototypes to evaluate: we will consider hands with several numbers of fingers and phalanges. Moreover, we will examine stimuli consisting in artificial arms with varying temporal sequences performing pointing and grasping tasks.

6 Conclusions

The present work extends the congruency effect successfully to non-human arms. Namely, a cartoon and robot-like hand elicited similar effects to human hands when observed by other humans. In addition, we showed that different arms/hands morphologies could have different effects. This result is important *per se:* it shows that artificial robotics agents affect people similarly. The second point is concerned with variations between morphologies: the trends we obtained should be explored more deeply to confirm that various morphologies affects observers sensory-motor system differentially. Last, we draw some research lines that could be of interest (guidelines) for designing companion robots easily acceptable by any user.

References

1. Mori, M.: Bukimi no tani The uncanny valley (K. F. MacDorman & T. Minato, Trans.). Energy 7(4), 33–35 (1970) (originally in Japanese)
2. Sternberg, S.: The discovery of processing stages: Extensions of Donders’ method. In: Koster, W.G. (ed.) Attention and Performance II. Acta Psychol., vol. 30, pp. 276–315 (1969)
3. Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W.: The theory of event coding (TEC): A framework for perception and action planning. Behav. Brain Sci. 24, 849–878 (2001)
4. Prinz, W.: Perception and action planning. Eur. J. Cogn. Psychol. 9, 129–154 (1997)
5. Hommel, B.: Grounding attention in action control: The intentional control of selection. In: Bruya, B.J. (ed.) Effortless Attention: A New Perspective in the Cognitive Science of Attention and Action, pp. 121–140. MIT Press, Cambridge (2010)
6. Wykowska, A., Schubö, A., Hommel, B.: How you move is what you see: Action planning biases selection in visual search. J. Exp. Psychol. Human 35, 1755–1769 (2009)

7. Wykowska, A., Hommel, B., Schubö, A.: Action-induced effects on perception depend neither on element-level nor on set-level similarity between stimulus and response sets. Atten. Percept. Psycho. 73, 1034–1041 (2011)
8. Schubotz, R.I., von Cramon, D.Y.: Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: An fMRI study. Neuroimage 15, 787–796 (2002)
9. Rizzolatti, G., Craighero, L.: The Mirror-Neuron System. Annu. Rev. Neurosci. 27, 169–192 (2004)
10. Rizzolatti, G., Fogassi, L., Gallese, V.: Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. 2, 661–670 (2001)
11. Gallese, V., Goldman, A.: Mirror neurons and the simulation theory of mind-reading. Trends Cogn. Sci. 2, 493–501 (1998)
12. Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J.C., Lenzi, G.L.: Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. P. Natl. Acad. Sci. USA 100, 5497–5502 (2003)
13. Fagioli, S., Hommel, B., Schubotz, R.I.: Intentional control of attention: Action planning primes action related stimulus dimensions. Psychol. Res. 71, 22–29 (2007)
14. Oberman, L.M., McCmeery, J.P., Ramachandran, V.S., Pineda, J.A.: EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots. Neurocomputing 70, 2194–2203 (2007)
15. Oztop, E., Franklin, D.W., Chaminade, T., Cheng, G.: Human–humanoid interaction: Is a humanoid robot perceived as a human? Int. J. Hum. Robot. 2, 537–559 (2005)
16. Chun, M.M., Wolfe, J.M.: Just Say No: How Are Visual Searches Terminated When There Is No Target Present? Cognitive Psychol. 30, 39–78 (1996)
17. Found, A., Müller, H.J.: Searching for unknown feature targets on more than one dimension: Investigating a ‘dimension weighting’ account. Percept. Psychophys. 58, 88–101 (1996)
18. Schubö, A., Schröger, E., Meinecke, C.: Texture segmentation and visual search for pop-out targets. Cognit. Brain Res. 21, 317–334 (2004)
19. Schubö, A., Wykowska, A., Müller, H.J.: Detecting pop-out targets in contexts of varying homogeneity: Investigating homogeneity coding with event-related brain potentials (ERPs). Brain Res. 1138, 136–147 (2007)
20. Matarić, M.J.: Socially assistive robotics. In: Coradeschi, S., et al. (eds.) Human-Inspired Robots (2006); IEEE Intelligent Systems, 81-83
21. Bartneck, C., Kanda, T., Ishiguro, H., Hagita, N.: Is the Uncanny Valley an Uncanny Cliff? In: Proceedings of the 16th IEEE, RO-MAN 2007, Jeju, Korea, pp. 368–373 (2007)
22. Brenton, H., Gillies, M., Ballin, D., Chatting, D.: The Uncanny Valley: does it exist? Paper Presented at HCI 2005, Animated Characters Interaction Workshop, Edinburgh, Napier University (2005)
23. MacDorman, K.: Subjective ratings of robot video clips for human-likeness, familiarity, and eeriness: An exploration of the uncanny valley. In: Proceedings of the ICCS/CogSci 2006 Long Symposium: Toward Social Mechanisms of Android Science, Vancouver, Canada, pp. 26–29 (July 2006)