

# Temporal Dynamics of Action Perception: The Role of Biological Appearance and Motion Kinematics

**Burcu Aysen Urgen** ([burgen@cogsci.ucsd.edu](mailto:burgen@cogsci.ucsd.edu))

Department of Cognitive Science, University of California, San Diego  
9500 Gilman Drive, La Jolla, CA 92093 USA

**Markus Plank** ([mplank@ucsd.edu](mailto:mplank@ucsd.edu))

Institute for Neural Computation, University of California, San Diego  
9500 Gilman Drive, La Jolla, CA 92093 USA

**Hiroshi Ishiguro** ([ishiguro@ams.eng.osaka-u.ac.jp](mailto:ishiguro@ams.eng.osaka-u.ac.jp))

Department of Adaptive Machine Systems, Osaka University  
Suita, Osaka, Japan

**Howard Poizner** ([hpoizner@ucsd.edu](mailto:hpoizner@ucsd.edu))

Institute for Neural Computation, University of California, San Diego  
9500 Gilman Drive, La Jolla, CA 92093 USA

**Ayse Pinar Saygin** ([saygin@cogsci.ucsd.edu](mailto:saygin@cogsci.ucsd.edu))

Department of Cognitive Science, University of California, San Diego  
9500 Gilman Drive, La Jolla, CA 92093 USA

## Abstract

We studied action perception and the role of visual form and visual motion kinematics of the observed agent using a stimulus set of human and humanoid robot actions and electroencephalogram (EEG). Participants viewed 2s. videos of three agents (Human, Android, Robot) performing recognizable actions: Human had biological form and motion, Android had biological form and non-biological motion, and Robot had non-biological form and non-biological motion. Early in processing (P200), Robot was distinguished from the other agents, likely due to low-level visual properties of the stimuli. We found a right temporal N170, which was most pronounced for Human, indicating possible modulation of this face- and body-sensitive ERP component by biological motion. There was a centro-parietal negativity (N300) that was most pronounced for Robot, and a later one (N400) for Human and Android. In the same time period (N300), Android was distinguished in the frontal channels from the other agents. A late positivity (P600) distinguished Human, again in frontal channels. These results highlight differential spatiotemporal cortical patterns during action perception depending on the viewed agent's form and motion kinematics.

**Keywords:** action perception; body perception; biological motion; social robotics; artificial agents; neuroimaging; EEG, ERP; uncanny valley

## Introduction

Successfully perceiving and understanding others' body movements is of biological significance, from hunting prey and avoiding predators, to communication and social interaction. The functional properties of the

neural systems that support action and body movement perception is currently an active research area in cognitive science and neuroscience.

Artificial agents such as robots can perform recognizable body movements, but can have varying degrees of biological appearance (form) and motion. As such, they provide us with an opportunity to study the specificity of neural responses to the seen agent's form and motion (as well as mismatches between the two). A prominent idea in action perception is *simulation theory*, whereby others' actions are understood via an internal sensorimotor simulation of the seen action in our own body representations (Barsalou, 2009). Supporting this, neural activity for action perception shows modulation by the degree of similarity between the observed action or actor, and the observers' own body (Buccino et al., 2004; Calvo-Merino, Grezes, Glaser, Passingham, & Haggard, 2006; Rizzolatti & Craighero, 2004). In terms of artificial agents such as robots, one might thus predict that increasing human-likeness engages simulation mechanisms more effectively.

On the other hand, human resemblance is not necessarily always a positive feature in artificial agent design. According to the *uncanny valley theory*, as an agent is made more human-like, the reaction to it becomes more and more positive and empathetic, until a point is reached at which the agent becomes oddly repulsive (Mori, 1970), an effect well-known in robotics and animation. Despite anecdotal evidence, there is little scientific data to characterize the uncanny valley (MacDorman & Ishiguro, 2006; Saygin,

Chaminade, Ishiguro, Driver, & Frith, 2011; Steckenfinger & Ghazanfar, 2009).

Previous studies on the perception of actions of humanoid robots have not found consistent results for or against simulation theory (Chaminade & Cheng, 2009). In a recent fMRI study, a more complex relationship between neural responses and the humanness of the observed agent was observed (including potential neural signals related to the uncanny valley), suggesting that focusing on simulation theory may be too narrow (Saygin, Chaminade, & Ishiguro, 2010). Furthermore, the specific role of biological appearance or biological motion in action processing have not been sufficiently explored in previous work, but is an area of interest in both social robotics and cognitive neuroscience (Chaminade, Hodgins, & Kawato, 2007; Kanda, Miyashita, Osada, Haikawa, & Ishiguro, 2008; Saygin, Chaminade, Urgen, & Ishiguro, 2011).

Although fMRI studies have identified the brain areas that are involved in action observation, much less is known about temporal aspects of body movement processing (Hirai, Fukushima, & Hiraki, 2003; Jokisch, Daum, Suchan, & Troje, 2005; Krakowski et al., 2011; Press, Cook, Blakemore, & Kilner, 2011). Since action processing is a naturally temporally unfolding event, it is important to further study its neural dynamics.

In the present study, we manipulated the form and the motion of the observed agent and recorded neural activity in the human brain using high-density electroencephalography (EEG), which allows us to investigate neurophysiological processes on a millisecond time scale. We used a unique stimulus set of well-matched human and humanoid robot actions (Saygin, Chaminade, Urgen et al., 2011). The stimuli consisted of videos of three agents: Human, Android, and Robot (Figure 1). Human had biological form and motion, Android had biological form and non-biological motion, and Robot had non-biological form and non-biological motion. The latter two were actually the same robot videotaped in two different appearances, but with identical kinematics. Another dimension of the stimuli was the congruence in the form and movement kinematics of the agents. Whereas Human and Robot had congruence in their form and movement kinematics (both being biological or non-biological, respectively), Android had incongruence in its form and movement kinematics as it had a biological appearance but non-biological movement kinematics.

Our goal is to study the temporal dynamics of action perception and its modulation by the seen agent's form and motion in relation to current theories in the field. Neural signals that may index simulation process would be expected to show some specificity to the Human condition. If the simulation process is driven primarily by appearance, responses to the Android are expected to be similar to the Human. If on the other

hand, biological motion is important for engaging simulation, Android responses are instead expected to show the same pattern as the Robot. As for the uncanny valley theory, we would expect neural responses for the Android to be distinct from the other conditions. Of course, the simulation theory and the uncanny valley theory are not mutually exclusive, and there may be evidence for both, possibly at different brain regions and in different time periods.

## Methods

### Participants

Twelve adults participated in the study. Participants were recruited from the student community at the University of California, San Diego (3 females, mean age: 24.4). All participants were right-handed, had normal or corrected-to-normal vision and no history of neurological disorders. Participants were either paid \$8 per hour or received course credit for their participation. They were informed about the nature of the study and signed consent forms in accordance with the UCSD Human Research Protections Program.

### Stimuli and Procedure

The experimental stimuli consisted of 2-second videos of three agents performing recognizable actions: A Human, an Android, and a Robot (Figure 1).

	ROBOT	ANDROID	HUMAN
Biological Motion	No	No	Yes
Biological Appearance	No	Yes	Yes
Congruent Motion and Appearance	Yes	No	Yes

Figure 1. Still frames from a drinking action for Robot, Android and Human agents and the experimental features of interest (form and motion).

The Android was Repliee Q2 (Ishiguro, 2006), and the Robot condition was the same robot in a modified appearance (Saygin, Chaminade, Ishiguro et al., 2011). We recorded EEG as participants watched video clips of the 3 agents carrying out five different upper body actions (drinking, picking an object, hand waving, talking, nudging). The experiment consisted of 15 blocks of 60 trials with equal number of videos of each agent.

The stimuli were displayed on a 22' Samsung monitor at 60 Hz. In order to prevent an augmented visual evoked potential at the beginning of the movie

onset that might occlude subtle effects between conditions, we displayed two consecutive gray screens (700-1000 ms and 500-700 ms, respectively) before each video clip. In order to minimize eye movement artifacts, subjects were instructed to fixate a fixation cross at the center of the screen. In order to control for subjects' attention throughout the experiment, every random 6-10 trials, a comprehension question was displayed (e.g., Drinking? Yes/No) and subjects responded with a bimanual key press.

### EEG Recordings and Analysis

EEG was recorded at 512 Hz from 64 ActiveTwo Ag/AgCl electrodes (Biosemi, Inc.) following the International 10-10 system. The electrode-offset level was kept below 25  $\mu$ V. Four additional electrodes were placed above and below the right eye, and lateral to the eyes to monitor oculomotor activity. The data were analyzed with MATLAB and the freely available EEGLAB toolbox (Delorme & Makeig, 2004). Data was high-pass filtered at 1 Hz, low-pass filtered at 50 Hz, and re-referenced to average mastoids. Atypical epochs of electromyographic activity were removed from further analysis by semiautomated epoch rejection procedures as implemented in EEGLAB. In order to discard eye-related artifacts, the data were decomposed by extended infomax ICA using binica as implemented in EEGLAB. The data were epoched time-locked to the onset of the video clips ranging from 200 ms preceding onset to 2000 ms after onset. Data was explored both qualitatively and quantitatively. Grand Average Event-related potentials (ERP) were computed using the BrainVision Analyzer 2 software package (BrainVision, Inc.). For display purposes, ERPs were low-pass filtered at 25 Hz.

Scalp topographies for the different conditions were generated. We identified specific channels and time periods for statistical analysis. For an unbiased analysis of differences between conditions, temporal regions of interest were determined from the mean grand average ERP activity across all conditions by visual inspection of all channels. The specific time window for each component was chosen to be the narrowest time window that was common to all channels that featured the respective component. This led to the selection of six time windows: 75-150 ms, 155-205 ms, 210-260 ms, 270-370 ms, 430-540 ms and 630-800 ms from stimulus onset. Not all channels had visible components in the ERP plots, but the temporal regions were chosen to be inclusive of all possible components of interest. Within each time window, we applied paired t-tests to compare individual mean amplitudes between conditions (Robot, Android, Human). The rationale of applying paired t-tests instead of ANOVA was because the former provide a test of our experimental hypotheses without considering

irrelevant comparisons. Since our design was not a full 2x2 factorial design with form and motion (lacking the non-biological form and biological motion condition) the main effect/interaction structure of a conventional ANOVA does not correspond to the experimental comparisons of interest (the effect of form, of motion, and of congruence of form and motion). Four of the analysed time windows showed the following 5 ERP components that significantly differed between experimental conditions: An occipital P200 (155-205 ms), a central temporal N170 (155-205 ms), a centro-parietal and frontal N300 (270-370 ms), a frontal N400 (430-540 ms), as well as a central and frontal P600 (630-800 ms). Where we presented data from selected channels, these were chosen as representative channels among those in the same region (as evident in the scalp distributions in Figure 2) for distinguishing one of the agents (Human, Android or Robot), thus showing the modulation of the respective component by form, motion, or congruence of form and motion. The reported p-values have been corrected for multiple comparisons unless stated otherwise (at alpha level 0.05).

### Results

EEG scalp topographies of the three conditions differed both spatially and temporally. Early on, the processing

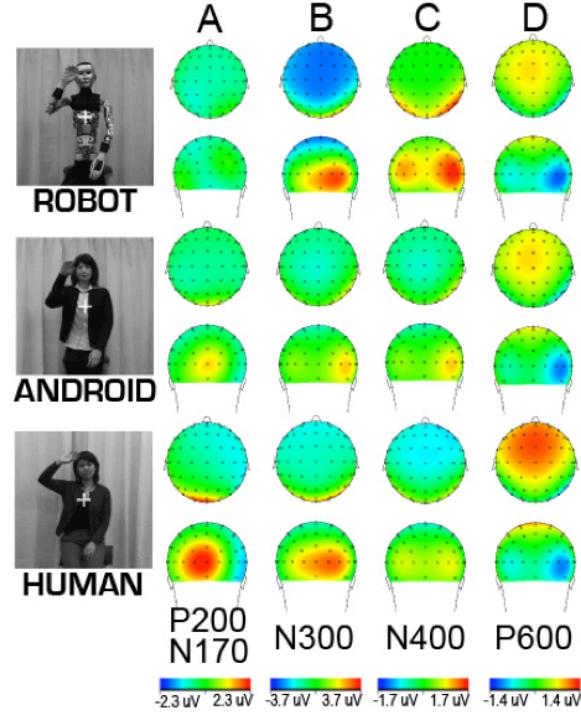


Figure 2. Scalp topographies corresponding to A) 155-205 ms (P200/N170), B) 270-370 ms (N300), C) 430-540 ms (N400), D) 630-800 ms (P600).

of Robot was distinguished from Human and Android, with an increased positivity across occipital regions for

the latter two agent conditions (Fig. 2A). Then, Robot was distinguished from Human and Android with a stronger negativity across frontal, central, and centro-parietal areas (Fig. 2B). Later, Robot was again distinguished from the other two agents with an increased positivity in centro-parietal regions, and Human was distinguished in the frontal regions (Fig. 2C). In a later stage, Human was distinguished with a stronger positivity in frontal regions (Fig. 2D).

The ERPs were then quantitatively compared across conditions to explore the role of biological form and biological action processing. Figure 3 shows ERP plots from representative channels in which the component of interest showed statistically significant amplitude modulations across conditions.

In the time window between 155-205 ms, we observed an occipital positivity (P200) that was stronger for Human and Android as compared to Robot ( $p<0.05$ ). Although Human elicited an increased P200 than Android, Human and Android did not differ significantly, indicating a form-based modulation of this component (Fig. 3, Iz). The same time window also showed an N170 in right centro-temporal channel T8, which showed a motion-sensitive amplitude modulation (Fig. 3, T8): Here, Human (featuring biological motion) elicited increased negative amplitude compared to Android and Robot ( $p<0.05$  and  $p<0.01$ , respectively); the latter conditions did not differ significantly.

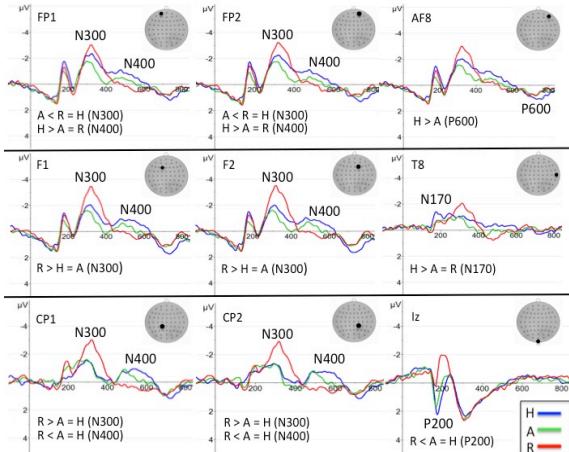


Figure 3. ERP Plots for selected channels depicting the condition effects for each component. (A: Android, H: Human, R: Robot)

Between 270 ms and 370 ms, there was a centro-parietal and frontal N300 (Fig. 3, CP1, CP2, F1, F2, Fp1, Fp2). Robot elicited a more pronounced negativity compared to Human and Android in frontal and centro-parietal channels bilaterally (Fig. 3 CP1:  $p<0.01$ ; CP2:  $p<0.01$ , F1:  $p<0.01$ ; F2:  $p<0.01$ ), indicating form-sensitive modulation. The N300 amplitudes of Human and Android did not differ. The same time window

showed less pronounced negativity for Android compared to Human and Robot in the most anterior-frontal channels bilaterally, possibly indicating a modulation by the (in)congruence of form and motion (Fig. 3, Fp1:  $p<0.001$ ; Fp2:  $p<0.05$ ). The responses for Human and Robot did not differ.

Between 430 ms and 540 ms, we observed a comparable negative amplitude in centro-parietal channels for Human and Android, which was absent for the Robot condition (Fig. 3 CP1, CP2), resulting in significant differences (Fig. 3 CP1 and CP2:  $p<0.05$ ). In the same time window, in frontal channels, Human elicited an increased negativity compared to Android and Robot (Fig 3. Fp1:  $p<0.05$ ; Fp2:  $p<0.01$ ).

Finally, between 630 ms and 800 ms, we observed a late positivity peaking in frontal channels, which was increased for Human vs. Android (Fig. 3 AF8:  $p<0.01$ ). The responses for Android and Robot did not differ in this time interval.

## Discussion

We investigated the temporal characteristics of neural activity during the perception of actions using a unique stimulus set of well-matched human and humanoid robot actions to manipulate the visual form and visual motion kinematics of the observed agent as we recorded electrical brain potentials (EEG). We found that neural activity during action perception is modulated differentially by the appearance and motion of the agent being observed, allowing us to observe the unfolding of perceptual and cognitive processes during action perception.

### P200

We found that an early stage of visual processing of the actions between 155-205 ms showed a form-sensitive modulation, where Robot (non-biological appearance) was distinguished from the other two agents (biological appearance, Figure 3 Iz). This is consistent with previous research on the P200 component, which is generally associated with early visual processing and is known to be sensitive to physical properties of visual stimuli (Luck & Hillyard, 1994). Since Robot had a distinct appearance from Human and Android, including low-level differences such as higher contrast and spatial frequencies, we interpret this effect as indicative of early perceptual differences, sensitive to the visual appearance of the agent being observed.

### N170

The early negative component N170, especially in the right hemisphere, has been associated with face and body processing in previous ERP research (de Gelder et al., 2010). In our study, the agents had different levels of anthropomorphism in their faces and bodies (i.e. biological vs. non-biological both in form and motion).

The Robot had a mechanical looking face with no movement, the Android and Human had similar facial appearance, but the Human face also featured biological motion (even though the actions used here did not feature prominent facial expressions and were upper body movements). We found that the amplitude of the N170 was modulated by the anthropomorphism of the agent, as manifested by a larger N170 for Human compared to the other agents. Since previous work on the N170 used static faces and bodies, our result may indicate that dynamic (biological) facial/bodily motion also elicits the N170. Another possibility is that the amplitude of the N170 might be differentially modulated depending on the presence of a context, as in our case the face was perceived together with the body during the performance of an action, whereas in previous work, still faces and bodies were shown as stimuli. As such, our results offer possible new studies to understand the functional significance of the N170 component.

### N300/N400 complex

The N300/N400 complex with an anterior distribution has been associated with the mapping of visual input onto representations in semantic memory (Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008). The increased centro-parietal negativity that we found for the robot condition in the 270-370 ms time interval (Figure 3, CP1 and CP2) over anterior regions may reflect a difficulty in mapping visual input onto existing semantic representations, since robots are currently not very familiar, certainly not in the context of actions such as those in our stimuli (e.g., drinking from a cup). If this interpretation is correct, we can also deduce this process being driven primarily by the form of the agent, for if motion was a factor, the Android were equally, if not more difficult to match to semantic memory. There was also a significant effect in the same time range in frontal channels, where Android differed from the other two agents (Figure 3, Fp1, Fp2). Given the Android represents a mismatch between form and motion being potentially linked to the uncanny valley phenomenon (Ishiguro, 2006; Saygin, Chaminade, Ishiguro et al., 2011), this could be a potential component to explore in future studies on the uncanny valley, or on congruence of form and motion more generally.

### P600 (late positivity)

In previous work, a late positivity or P600 has mostly been studied in the domain of language and is most commonly associated with syntactic processing (Friederici, 2004). Few studies have interpreted the P600 in other domains (Sitnikova et al., 2008). In our data (Figure 3, AF8), we found that this component was elicited most strongly by the Human condition. This can lead us new avenues of research to understand the

functional significance of the ERP components observed in action perception.

### Implications for Action Processing

Although action processing has been an active area of study in cognitive neuroscience, most work to date has used fMRI rather than electrophysiology. More specifically, a number of studies have focused on the perception of human and robot agents with fMRI, with inconsistent support for the simulation theory (Saygin, Chaminade, Ishiguro et al., 2011). Here, we add new ERP results to this literature, providing information about the role of humanoid form and humanoid motion during the course of action perception.

The stimuli used here were previously utilized in an fMRI repetition-suppression study in which brain activity did not show evidence for form-based or motion-based simulation *per se*, but instead was most significantly affected by form-motion incongruence (Saygin, Chaminade, Ishiguro et al., 2011). Here, with a more time-resolved method, we found distinct stages of processing during which neural responses differed based on both the form and the motion of the seen agent. These effects were likely lost due to the temporal insensitivity of fMRI, highlighting the importance of using multiple, complementary techniques.

A well-known face-sensitive component, the N170, was elicited by our stimuli. Previous work on this ERP signature of face processing has used static face stimuli, as opposed to movies including the body as we did here. Our data suggest new possible ways in which the N170 can be modulated. Specifically we hypothesize that either biological motion of the face and/or the context provided by the body are modulators of the N170.

Our data did not reveal patterns of activity that can be linked straightforwardly to simulation theory. There was some selectivity for the Human (for whom simulation theory would predict differential effects, whether driven by form or motion) for the frontal N400 and P600, but there is little prior literature on actions for these components, and no link to sensorimotor simulation that we are aware of. The uncanny valley theory also cannot account for all of the patterns in our data, although the frontal N300 response could be interpreted as biomarker for the uncanny valley. These components should be viewed as possible indices related to each theory, to be tested in new studies.

Overall, in this first ERP study of action perception with human and humanoid agents, we highlight the complexity of action processing that can be revealed using more time-resolved methods. We found distinct neural signatures of the viewed agent's form and motion in different time periods, both early (perceptual) and late (cognitive) in processing. These results do not globally fit into either simulation or uncanny valley frameworks, although a focus on specific components

such as the N170 and N300/400 in upcoming studies might help better understand the mechanisms of action perception and its neural basis. Work on neural dynamics of action processing can not only shed light on the cognitive neuroscience of action perception, but also to inform the burgeoning field of social robotics (Saygin, Chaminade, Urgen et al., 2011).

### Acknowledgements

This research was supported by the Kavli Institute for Brain and Mind (APS), California Institute of Telecommunications and Information Technology (Calit2) (BAU, APS), NSF (SBE-0542013, Temporal Dynamics of Learning Center), and ONR (MURI Award # N00014-10-1-0072, HP). The authors would like to thank Marta Kutas, Arthur Vigil, members of the Intelligent Robotics Laboratory (Osaka University) for the creation of the stimuli, and Joe Snider (Poizner Lab) for his help in the experimental setup.

### References

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London B*, 364(1521), 1281-1289.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., et al. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an fMRI study. *Journal of Cognitive Neuroscience*, 16(1), 114-126.

Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, 16(19), 1905-1910.

Chaminade, T., & Cheng, G. (2009). Social cognitive neuroscience and humanoid robotics. *Journal of Physiology Paris*, 103(3-5), 286-295.

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, 2(3), 206-216.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9-21.

de Gelder, B., Van den Stock, J., Meeren, H.K.M, Sinke, J.B.A., Kret, M.E., Tamietto, M. (2010). Standing up for the body: Recent progress in uncovering the networks involved in processing bodies and bodily expressions. *Neurosci. Biobehav. Rev.*, 34, 513-527.

Friederici, A. D. (2004). Event-related brain potential studies in language. *Current Neurology Neuroscience Report*, 4(6), 466-470.

Hirai, M., Fukushima, H., & Hiraki, K. (2003). An event-related potentials study of biological motion perception in humans. *Neurosci Lett*, 344(1), 41-44.

Ishiguro, H. (2006). Android science: conscious and subconscious recognition. *Connection Science*, 18(4), 319-332.

Jokisch, D., Daum, I., Suchan, B., & Troje, N. F. (2005). Structural encoding and recognition of biological motion: evidence from event-related potentials and source analysis. *Behavioral Brain Research*, 157(2), 195-204.

Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., & Ishiguro, H. (2008). Analysis of humanoid appearances in human-robot interaction. *IEEE Transactions on Robotics*, 24(3), 725-735.

Krakowski, A. I., Ross, L. A., Snyder, A. C., Sehatpour, P., Kelly, S. P., & Foxe, J. J. (2011). The neurophysiology of human biological motion processing: A high-density electrical mapping study. *Neuroimage*, 56(1), 373-383.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31(3), 291-308.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3), 297-337.

Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33-35.

Press, C., Cook, J., Blakemore, S. J., & Kilner, J. M. (2011). Dynamic modulation of human motor activity when observing actions. *Journal of Neuroscience*, 31(8), 2792-2800.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-192.

Saygin, A. P., Chaminade, T., & Ishiguro, H. (2010). The perception of humans and robots: Uncanny hills in parietal cortex. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2716-2720). Portland, OR: Cognitive Science Society.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. F. (2011). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*.

Saygin, A. P., Chaminade, T., Urgen, B. A., & Ishiguro, H. (2011). Cognitive neuroscience and robotics: A mutually beneficial joining of forces In L. Takayama (Ed.), *Robotics: Systems and Science*. Los Angeles, CA.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, 20(11), 2037-2057.

Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences of the United States of America*, 106(43), 18362-18366.