Disappearance of Inversion Effect for Walking Animation with Robotic Appearance

Masahiro Hirai (hirai@ardbeg.c.u-tokyo.ac.jp)

Department of General System Studies,
The University of Tokyo,
3-4-1 Komaba, Meguro-ku, Tokyo, 153-8902 JAPAN

Kazuo Hiraki (khiraki@idea.c.u-tokyo.ac.jp)

Department of General System Studies,
The University of Tokyo,
3-4-1 Komaba, Meguro-ku, Tokyo, 153-8902 JAPAN

Abstract

Recent studies have reported similarity in the neural processing of human and robot actions; however, whether this is the case remains controversial. Here, we examined this controversy using the inversion effect, a phenomenon whereby an upright face- and body-sensitive event-related potential component is enhanced and delayed in response to an inverted face and body, but not an inverted object. The results showed that the inversion effect occurs only with a human, not with robotic and point-light appearances, suggesting that our visual system differentially processes human and robot actions.

Introduction

It has been suggested that our neural system is tuned specifically to be able to detect the human body. For example, a previous psychophysical study revealed that the neural system differentially processes the human body and objects (Shiffrar & Freyd, 1990), while recent neuroimaging studies have shown specific tuning to the human body (Downing et al., 2001), the human face (Gauthier et al., 2000; Kanwisher, 2000) and human body movements (e.g. biological motion; Grossman et al., 2000). These findings imply that our neural system responds sensitively to both human appearance and motion.

With the recent development of robotic technologies, living with robots has become a reality, not just something seen in science fiction movies. Moreover, various kinds of robots now appear in our daily lives; for example, humanlike robots (Collins et al., 2005) such as "ASIMO" [1], "QRIO"[2] and "Robovie" (Ishiguro et al., 2003), which were designed and developed specifically for household use. Furthermore, robots with a very human-like appearance are now being developed, and at a glance, are often indistinguishable from human beings [3].

Since these robots have similar appearance information to humans, such as body structure and configuration, yet are not a biological object, the question therefore arises as to whether or not our neural system interprets such robots as a kind of human. To date, several studies have provided clues to answer this question. For example, in a behavioral study, Kilner et al. (2003) reported that observation of other humans making incongruent movements, but not robots, had a significant interference effect on executed movements of participants. On the other hand, Pelphrey et al. (2003) reported that activation of the superior temporal sulcus (STS) during processing of the human appearance is similar to that during processing of a robotic appearance. The former study suggests that superficial information might affect our perception-action system, while the latter suggests that motion information, not just superficial information, might also affect activation of the STS.

Intuitively, both appearance and motion information therefore seem to play an important role in detecting characteristics of 'human-likeness'. That is, our visual system discriminates humans from objects not only by detecting appearance information, but also using motion information such as biological motion perception (Johansson, 1973). However, the relationship between appearance and motion information in detecting 'human-likeness' has not been fully investigated.

The aim of the present study is to clarify how different appearance information with identical motion information affects the neural response. To investigate this, we recorded event-related potentials (ERPs) in human participants and evaluated the occurrence of the inversion effect (Bentin et al., 1996; Linkenkaer-Hansen et al., 1998; Rossion et al., 2000; Taylor et al., 2001; Itier & Taylor, 2004; Stekelenburg & de Gelder, 2004). The inversion effect is a phenomenon whereby an upright face- and body-sensitive ERP component (N170) is delayed (Bentin et al., 1996; Linkenkaer-Hansen et al., 1998; Taylor et al., 2001) and enlarged in amplitude (Linkenkaer-Hansen et al., 1998; Rossion et al., 2000; Taylor et al., 2001; Itier & Taylor, 2004) in response to inverted faces and bodies but not inverted objects (Rossion et al., 2000; Stekelenburg & de Gelder, 2004). An inversion effect has also been reported in magnetoencephalography (MEG) (Watanabe et al., 2003) and functional magnetic resonance imaging (fMRI) (Haxby et al., 1999) studies of upright and inverted face perception.

In this study, we employed three kinds of walking animation with different superficial information (human, robot and point-light appearance) to explore two hypotheses. The first hypothesis is that if robotic walking animation is processed like an object, the inversion effect will not occur. However, in contrast, if it is processed like human information, the inversion effect will be observed as in the human appearance condition. The second hypothesis

---

[1] http://asimo.honda.com/
[2] http://www.sony.net/SonyInfo/QRIO/
[3] http://news.bbc.co.uk/1/hi/sci/tech/4714135.stm

is that if superficial information does not affect processing of human walking animation, ERP waveforms in all three conditions will show similar patterns because of the identical walking actions. Both hypotheses were tested by measuring ERPs.

Materials and Methods

Three kinds of walking animation (Fig. 1) (human, robot and point-light) with two orientations (upright and inverted) were employed. The structure of the body and walking speed were identical in all animations. Nineteen healthy participants were included as study participants. They were required to view each animation passively and mentally count the number of asterisks appearing randomly during each block. Electroencephalograms (EEGs) were recorded during each trial with a Geodesic Sensor Net composed of 64 electrodes (Tucker, 1993).

Participants

We studied nineteen healthy participants (range/mean age: 18-30/23.7 ± 3.9 years; 14 males, 5 females). Seventeen subjects were right-handed and all had normal or corrected-to-normal vision. All subjects provided informed consent for a protocol that was approved by the Ethics Committee of the University of Tokyo.

Experimental Procedure

Six experimental conditions were employed as shown in Fig.1. To generate the animated figures, we used the Poser 5.0 software program (Curious Labs, Santa Cruz, CA). Both the human and robot animations were generated using built-in 3D models. For the point-light animation, the human 3D model used in the human animation was replaced by 14 small balls placed at all joints and the head using Metasequoia (Mizuno Lab, Japan).

All animations were viewed in profile as walking as if on a treadmill. The walking speed in all animations was 2.0 steps per second. The animations were displayed on a 17-inch monitor against a black background. Each participant was seated 100 cm from the display in a dimly lit room. The entire visual stimulus was approximately 3 × 3°. To produce smooth animated motion, each animation comprised 15 frames displayed for 510 ms and with an interframe interval of approximately 34 ms. The initial number of frames was randomized to prevent the participants from remembering the initial starting figure.

Each experiment consisted of eight blocks with a 1 min inter-block interval. Twenty stimuli were employed in each animation condition, and accordingly, 120 animations were presented per block and 960 per experiment. Thus, each animation was presented 160 times throughout each experiment. In each trial, the stimulus was presented for 510 ms followed by presentation of a white fixation point (a 0.3 × 0.3° cross) for 500 ms. To ensure that subjects maintained their gaze on the center of the monitor during all animations, participants were asked to engage in a continuous performance task. They were asked to count the number of times a yellow asterisk appeared randomly on the screen and
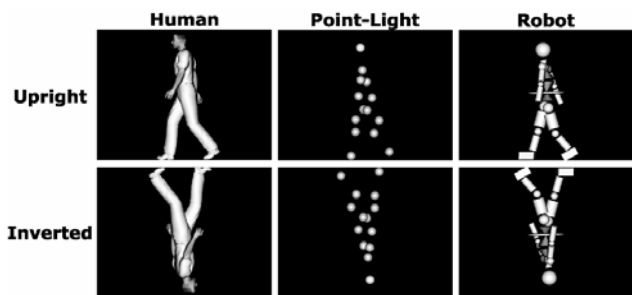

Figure 1: The experimental stimuli. All animations were viewed in profile as walking as if on a treadmill.

report back at the end of each block. The asterisk was presented for 500ms instead of the walking animation eight times per block.

Results

Behavior results

The percentage of correct performance in the counting task was 46.1 ± 24.3 % (average ± S.D).

ERP results

Fig.2 shows the grand mean waveforms of the ERP responses. As in our previous study (Hirai et al., 2005), we collapsed the three electrodes surrounding each T5/T6 (International 10-20 System) into two sites. A single negative peak was found at around 200 ms (conventional N170-like component) in both the human and robot conditions, while in the point-light condition two negative peaks were observed at 200 and 340 ms, respectively. The peak latency and amplitude (in order to correct the N1 amplitude, we calculated the P1-N1 amplitude) of each component were also calculated, and subsequently, statistical analysis was carried out.

P1-N1 amplitude and N1 latency In three-way ANOVA of the P1-N1 amplitude, laterality × type of appearance × orientation was significant [$F(2,36) = 3.37$, $p<0.05$]. Subsequent analysis revealed that the amplitude in the right hemisphere was significantly larger with the inverted orientation than the upright orientation in the human appearance condition [4.24μV vs. 5.18μV, $F(1,108) = 9.62$, $p<0.01$]. In addition, the amplitude in the left hemisphere was significantly larger than that in the right hemisphere in the upright-human condition [5.09μV vs. 4.24μV, $F(1,108) = 4.14$, $p<0.05$]. Moreover, the amplitude with the human condition was significantly larger than that with the point-light condition in the left hemisphere with the upright orientation [5.09μV vs. 3.81μV, $p<0.01$; Tukey's HSD]. The amplitude with the human condition was also significantly larger than that with the point-light condition in both hemispheres with the inverted orientation [left hemisphere: 5.29μV vs. 4.24μV, $p<0.05$; right hemisphere: 5.18μV vs. 3.89μV, $p<0.01$; Tukey's HSD]. Similarly, the amplitude with the human condition was significantly larger than that with the robot condition in both hemispheres with the inverted orientation [left hemisphere: 5.29μV vs. 4.39μV,
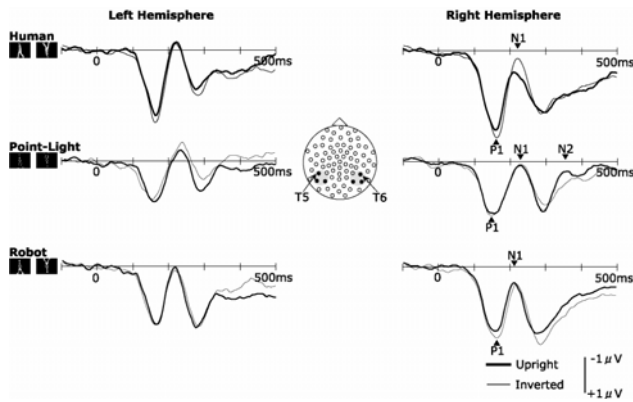
Figure 2: Grand averaged ERP waveforms showing the inversion effect with each appearance and orientation condition.

$p<0.05$; right hemisphere: 5.18μV vs. 4.13μV, $p<0.05$; Tukey's HSD]. The main effect of orientation was also significant with the N1 latency, $[F(1,18) = 4.72, p<0.05$; upright: 218.6ms vs. inverted: 224.0ms], indicating that the latency of inverted stimuli was longer than that of upright stimuli.

P1-N2 amplitude and N2 latency The N2 component was observed only with the point-light condition, and accordingly, the P1-N2 amplitude and N2 latency were analyzed as above. Two-way of ANOVA was applied to the P1-N2 amplitude using laterality (left or right) and orientation (upright or inverted) as variables. As a result, laterality × orientation was shown to be significant $[F(1,18) = 10.9, p<0.01]$. Subsequent analysis revealed that the amplitude with the inverted condition was significantly larger than that with the upright condition in the left hemisphere [4.03μV vs. 3.48μV, $F(1,36) = 6.28, p<0.01]$. On the contrary, the amplitude with the upright condition was significantly larger than that with the inverted condition in the right hemisphere [3.68μV vs. 3.01μV, $F(1,36) = 9.74, p<0.01]$. Moreover, the amplitude in the left hemisphere was significantly larger than that in the right hemisphere with the inverted condition [4.03μV vs. 3.01μV, F(1,36)=5.53, $p<0.01]$. No significance was observed regarding the N2 latency in the point-light motion condition.

The P1-N1 amplitude in the right hemisphere was significantly larger with the inverted orientation than the upright orientation with the human appearance condition [4.24μV vs. 5.18μV, $F(1,108) = 9.62, p<0.01]$. This was not observed with the other appearance conditions. Regarding the N1 latency, the main effect of orientation was also significant $[F(1,18) = 4.72, p<0.05$; upright: 218.6ms vs. inverted: 224.0ms], indicating that the latency of the inverted stimuli was longer than that of the upright stimulus.

Conclusion and Discussion

Our data demonstrated that the inversion effect occurs in the right occipitotemporal region with the human appearance condition only. These findings are consistent with the results of recent neuroimaging studies of face and body perception (Bentin et al., 1996; Linkenkaer-Hansen et al., 1998; Rossion et al., 2000; Taylor et al., 2001; Itier & Taylor, 2004; Stekelenburg & de Gelder, 2004). With regard to latency, a recent study suggested that the delay in latency of the N1 component is observed not only with faces but also objects (Itier et al., 2006), which is also consistent with our present data. The present findings imply that robot walking animation is not processed like human information (i.e. robots are not categorized as humans), even though the robots are analogous in appearance and have identical motion properties (speed and motion trajectory). This suggests that appearance information affects the neural responses and this categorization is processed within early visual processing.

In the light of recent findings, our data seems inconsistent with those of Pelphrey et al. (2003) who suggests the importance of motion information only. That is, our results show that appearance information also has an affect on the neural response in the occipitotemporal region, which might be involved with the STS region (Homan et al., 1987). However, for the following two reasons, we believe these findings are in fact consistent. First, Pelphrey et al. (2003) used an fMRI technique to investigate conditional differences, and thus, could not measure the neural response with millisecond temporal resolution. On the other hand, in our study, the conditional differences were observed at around 200ms after stimulus onset, a rapid response that neuroimaging techniques such as fMRI are perhaps unable to detect. The second reason is related to their use of upright and scrambled conditions only; that is, they did not investigate the inversion effect. Consistent with their fMRI findings, our ERP data showed no conditional difference in the P1-N1 amplitude between the upright-robotic and upright-human conditions.

As in our previous study, we found two negative peaks in the occipitotemporal region at around 200 and 340ms, respectively, with the point-light motion condition. The second negative component is thought to reflect processing of biological motion (Hirai et al., 2003, 2005), specific analysis of motion patterns providing biologically relevant information (Jokisch et al., 2005) or form-from-motion processing (Wang et al., 1999).

The conditional differences in the inversion effect in the present study might also be explained from the point of view of perceptual expertise. Several studies have reported that for such experts (e.g. dog show judges) processing of a car, dog, or bird is similar to processing of a face (Gauthier & Tarr, 2002; Diamond & Carey, 1986). Another ERP study also showed inversion of the N170 component in response to a non-face object (Greebles) as well as faces with expertise training in Greebles (Rossion et al., 2002). Accordingly, it is likely that an object with a robotic appearance is not observed frequently, unlike the human body, and thus, this frequency of contact might have elicited the conditional difference in the inversion effect.

In conclusion, using the inversion effect as an index, this study clarified that a human walking appearance is processed differently from robotic and point-light walking appearances. These findings indicate that our visual system distinctly and differentially processes biological

appearances such as the human body and nonbiological appearances such as robots within a short latency, even when the motion information is superimposed. To determine the role of motion information, further work is needed to fully elucidate the differential neural responses to robotic and human motion with identical appearance information.

Acknowledgments

References

Bentin, S., Allison, T., Puce, A., Perez, A., McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8,* 551-565.

Collins, S., Ruina, A., Tedrake, R., Wisse, M. (2005). Efficient bipedal robots based on passive-dynamic walkers. *Science,* 307, 1082-1085.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General,* 115, 107-117.

Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science,* 293, 2470-2473.

Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3,* 191-197.

Gauthier, I., & Tarr, M.J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance,* 28, 431-446.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan V., Neighbor, G., Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, 12, 711-720.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L., Hoffman, E.A., Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron,* 22, 189-199.

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

Hirai, M., Senju, A., Fukushima, H., Hiraki, K. (2005). Active processing of biological motion perception: an ERP study. *Brain Research: Cognitive Brain Research,* 23, 387-396.

Homan, R.W., Herman, J., Purdy, P. (1987). Cerebral location of international 10-20 system electrode placement. *Electroencephalography and Clinical Neurophysiology*, 66, 376-382.

Ishiguro, H., Ono, T., Imai, M., Kanda, T. (2003). Development of an interactive humanoid robot "Robovie" -An interdisciplinary approach, R. A. Jarvis and A. Zelinsky (Eds.), *Robotics Research,* Springer, pp. 179-191.

Kilner, J.M., Paulignan, Y., Blakemore, S.J. (2003). An interference effect of observed biological movement on action. *Current Biology,* 13, 522-525.

Itier, R.J., & Taylor, M.J. (2004). Source analysis of the N170 to faces and objects. *Neuroreport*, 15, 1261-1265.

Itier, R.J., Latinus, M., Taylor, M.J. (2006). Face, eye and object early processing: what is the face specificity? *Neuroimage,* 29, 667-676.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics,* 14, 201-211.

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, 195-204.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3, 759-763.

Linkenkaer-Hansen, K., Palva, J.M., Sams, M., Hietanen, J.K., Aronen, H.J., Ilmoniemi, R.J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters,* 253, 147-150.

Pelphrey, K.A., Mitchell, T.V., McKeown, M.J., Goldstein, J., Allison, T., McCarthy, G. (2003). Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *Journal of Neuroscience*, 23, 6819-6825.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M.J., Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science,* 13, 250-257.

Rossion, B., Gauthier, I., Tarr, M.J., Despland, P., Bruyer, R., Linotte, S., Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, 11, 69-74.

Shiffrar, M., & Freyd, J.J. (1990). Apparent motion of the human body. *Psychological Science,* 1, 257-264.

Stekelenburg, J.J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. *Neuroreport*, 15, 777-780.

Taylor, M.J., Edmonds, G.E., McCarthy, G., Allison T. (2001). Eyes first! Eye processing develops before face processing in children. *Neuroreport,* 12, 1671-1676.

Tucker, D.M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, 87, 154-163.

Wang, J., Jin, Y., Xiao, F., Fan, S., Chen, L. (1999). Attention-Sensitive Event-Related Potentials Elicited by Kinetic Forms. *Clinical Neurophysiology*, 110, 329-341.

Watanabe, S., Kakigi, R., Puce, (2003). A. The spatiotemporal dynamics of the face inversion effect: a magneto- and electro-encephalographic study. *Neuroscience,* 116, 879-895.