Vision Research 135 (2017) 10–15

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Action recognition is viewpoint-dependent in the visual periphery

Laura Fademrecht *, Isabelle Bülthoff, Stephan de la Rosa

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

ARTICLE INFO

Article history:
Received 4 November 2016
Received in revised form 13 January 2017
Accepted 14 January 2017
Available online 26 April 2017

Number of reviews = 2

Keywords:
Action recognition
Peripheral vision
Viewpoint-dependency
Action categorization

ABSTRACT

Recognizing actions of others across the whole visual field is required for social interactions. In a previous study, we have shown that recognition is very good even when life-size avatars who were facing the observer carried out actions (e.g. waving) and were presented very far away from the fovea (Fademrecht, Bülthoff, & de la Rosa, 2016). We explored the possibility whether this remarkable performance was owed to life-size avatars facing the observer, which – according to some social cognitive theories (e.g. Schilbach et al., 2013) – could potentially activate different social perceptual processes as profile facing avatars. Participants therefore viewed a life-size stick figure avatar that carried out motion-captured social actions (greeting actions: handshake, hugging, waving; attacking actions: slapping, punching and kicking) in frontal and profile view. Participants' task was to identify the actions as 'greeting' or as 'attack' or to assess the emotional valence of the actions. While recognition accuracy for frontal and profile views did not differ, reaction times were significantly faster in general for profile views (i.e. the moving avatar was seen profile on) than for frontal views (i.e. the action was directed toward the observer). Our results suggest that the remarkable well action recognition performance in the visual periphery was not owed to a more socially engaging front facing view. Although action recognition seems to depend on viewpoint, action recognition in general remains remarkable accurate even far into the visual periphery.

1. Introduction

Most of the actions that we encounter in everyday life are likely to fall within the visual periphery. Depending on the social engagement, the viewpoint of those action changes. If the actions are directed towards the observer, as it is the case in social interactions, chances are that an observer sees the action from a front facing perspective. In contrast, if the observer is a third person not involved in the interaction, it is likely that she will see the actions from a different viewpoint, e.g. profile view. Does action recognition performance in the visual periphery depend on the viewpoint of the action?

1.1. Viewpoint sensitivity

Viewpoint-specific encoding of visual information by neural units seems to be a general organizational principle of the visual system (for an alternative view see Biederman & Gerhardstein, 1993). There is evidence that visual processes involved in the recognition of objects (Logothetis & Pauls, 1995; Logothetis, 1995; Logothetis, Pauls, Bülthoff, & Poggio, 1994; Tarr, 1995; Tarr & Bülthoff, 1998), faces (Bruce, Valentine, & Baddeley, 1987; Hill, Schyns, & Akamatsu, 1997; Perrett, Hietanen, Oram, & Benson, 1992; Troje & Bülthoff, 1996; Troje & Kersten, 1999) and actions (Daems & Verfaillie, 1999; de la Rosa, Mieskes, Bülthoff, & Curio, 2013; Jokisch, Daum, & Troje, 2006; Troje et al., 2005; Verfaillie, 1993) are sensitive to the viewpoint of the stimulus. Physiological single cell studies provide supporting evidence by showing that some cells in the temporal cortex are only activated by a particular viewpoint of objects, faces or bodies (Barraclough, Keith, Xiao, Oram, & Perrett, 2009; Jellema, Maassen, & Perrett, 2004; Jellema & Perrett, 2003, 2006; Logothetis et al., 1994; Perrett et al., 1989, 1992). As for actions, viewpoint dependent recognition effects have been reliably found under varying testing conditions. For example, viewpoint dependent recognition has been found using biological motion stimuli in a recognition task (de la Rosa et al., 2013; Jokisch et al., 2006), in an identification task (Prasad & Shiffrar, 2009; Troje et al., 2005), as well as in an adaptation paradigm with computer-generated mannequins (Benton, Thirkettle, & Scott-Samuel, 2016). However, the viewpoint dependency of action recognition seems to depend on whether the actions are carried out by oneself or by others (Jokisch et al., 2006; Prasad & Shiffrar, 2009; Troje et al., 2005). Moreover, Daems and Verfaillie

* Corresponding author.
E-mail address: laura.fademrecht@tuebingen.mpg.de (L. Fademrecht).

http://dx.doi.org/10.1016/j.visres.2017.01.011
0042-6989/© 2017 The Author(s). Published by Elsevier Ltd.

(1999) showed that priming stimuli that had the same orientation as the test stimuli were more effective than their mirror-images in an action naming task. Verfaillie (1993) examined the effects of orientation of a point-light walker (i.e. walking to the left or to the right) using short-term priming. Subjects discriminated between a point-light walker and a nonhuman walker. Their results revealed that priming effects only occurred when the priming walker and the test walker had the same orientation. Additionally, physiologically grounded computational models of action recognition outline how viewpoint-sensitive action recognition units are linked to behavioral performance. These models show that view-dependent action recognition can be explained in a physiologically plausible way (Fleischer, Caggiano, Thier, & Giese, 2013; Giese & Poggio, 2003; Lange & Lappe, 2006). Overall, there is strong evidence in favor of the idea that action recognition mechanisms are tuned to specific views.

1.2. Preferred viewpoints for action recognition

The functional role of viewpoint-dependent action recognition mechanisms is unknown. It is possible that viewpoint dependent action recognition is tied to the active social involvement in an interaction in the sense that active engagement in a social interaction often causes observers to see an action from a frontal view. In contrast, observing other people will result in mainly non-frontal views, e.g. profile views. In line with this idea some social cognitive theories suggests the primacy of first person over other views (Vogeley & Fink, 2003). The origin of this effect is believed to be the activation of perceptual cognitive processes under first person viewing conditions that resemble those when participants are actively engaged in a social interaction (Schilbach et al., 2013). Supporting evidence for this first person perspective comes from studies showing that participants give socially relevant facial expressions a higher rating when they are directed toward them than toward a third person (Schilbach et al., 2006). Furthermore, neural activation patterns differed when the facial expressions were directed towards the observer or not. Hence one might expect that front facing actions are better recognized than profile views of the same actions due to their larger visual resemblance to real social interactions.

1.3. Differences in visual action recognition between the fovea and the visual periphery

Some evidence suggests that the viewpoint dependent encoding of actions as observed in foveal vision might not straightforwardly apply to peripheral action recognition. Existing studies concerning the perception of biological motion have mainly focused on detection and direction discrimination of locomotive actions (e.g., walking, running) at eccentricities up to 12° (near periphery). Their results show that these actions can be readily detected at small eccentricities (up to 12°), although there was always a disadvantage in the periphery compared with central vision (Ikeda, Blake, & Watanabe, 2005; Ikeda, Watanabe, & Cavanagh, 2013; Thompson, Hansen, Hess, & Troje, 2007). Thompson et al. (2007) provide evidence that compared to foveal vision the visual periphery suffers from a deficit in segregating signal from noise. Thurman and Lu (2013) showed that in peripheral vision local motion cues, orientation cues and spatial cues interact with each other whereas foveal vision is dominated by global motion cues. These differences between foveal and peripheral vision leave open the question whether orientation sensitivity for action recognition differs between central and peripheral vision. In the present study, one aim was to investigate this question. In the current study, we wanted to examine viewpoint dependent recognition of actions in the visual periphery. Since it is well known that different tasks are associated with different action recognition performances (de la Rosa et al., 2014; Fademrecht, Bülthoff, & de la Rosa, 2016). We used two recognition tasks to investigate the influence of action orientation (first and third perspective) and action presentation (foveal vs peripheral) on participants' action recognition performance. Specifically in the first recognition task participants categorized actions. Specifically, participants saw six actions (i.e. shaking hands, hugging, waving, slapping, punching and kicking) and reported whether they saw a greeting or an attack. In the other task, participants were asked to report the emotional valence of the same actions (valence task).

2. Methods

2.1. Participants

30 participants (11 males, 20 females) from the local community of Tübingen participated in the experiment. The age ranged from 21 to 32 years (mean: 25.5). All participants received monetary compensation for their participation and gave their informed written consent prior to the experiment. The participants had normal or corrected to normal vision (using contact lenses). The study was conducted in accordance with the Max Planck Society policy and the Code of Ethics of the World Medical Association and has been approved by the ethics committee of the University of Tübingen.

2.2. Stimuli

The same stimuli were used as in Fademrecht et al. (2016). Six actions were acted out by six actors (three female) and recorded via motion capture. Three actions with positive emotional valence (handshake, hugging and waving) and three actions with negative valence (slapping, punching and kicking) were acted out by each actor six times, leading to 216 stimuli in total. The action sequences used as stimuli lasted between 800 and 1500 ms. Each action started with the actor standing in a neutral position and ended with the peak frame of the action. The peak frame of an action is defined here as the point in time just before the actor started moving back into the neutral position. The motion data was mapped onto a grey life-size 'stick-figure avatar' (avatar height: 170 cm, about 32° visual angle; see Fig. 1) that participants viewed on a large screen (see below for more details). The figures were either oriented towards the participant (frontal view, first person perspective) or orthogonal to the participant's direction of view (profile view, third person perspective). A stick figure was used instead of a full-fleshed avatar to prevent any other visual cues apart from motion from influencing participant's recognition judgements (for more details about the stimuli, see Fademrecht et al. (2016)).

2.3. Apparatus

Stimuli were presented in 2D on a large panoramic screen with a semi-cylindrical projection system. The semi-circular wide screen was 7 m long (diameter) and 3.2 m high (230° horizontally, 125° vertically). Six EYEVIS LED DLP projectors (1920×1200, 60 Hz) were used to display the stimuli against a grey background. The geometry of the screen can be described as a quarter-sphere. The visual distortions caused by the curved projection screen were compensated with the use of warping technology software. With this setup visual stimuli can be presented to the whole horizontal human visual field. Participants placed their head on a chin and forehead rest. An eye tracker (Eyelink II, SR Research Ltd., Canada) was used to control for eye movements. If the participant's gaze