Exp Brain Res
DOI 10.1007/s00221-007-1038-5

RESEARCH NOTE

Initial fixation placement in face images is driven by top–down guidance

Kun Guo

Received: 17 May 2007 / Accepted: 15 June 2007

Abstract The eyes are often inspected first and for longer period during face exploration. To examine whether this saliency of the eye region at the early stage of face inspection is attributed to its local structure properties or to the knowledge of its essence in facial communication, in this study we investigated the pattern of eye movements produced by rhesus monkeys (*Macaca mulatta*) as they free viewed images of monkey faces. Eye positions were recorded accurately using implanted eye coils, while images of original faces, faces with scrambled eyes, and scrambled faces except for the eyes were presented on a computer screen. The eye region in the scrambled faces attracted the same proportion of viewing time and fixations as it did in the original faces, even the scrambled eyes attracted substantial proportion of viewing time and fixations. Furthermore, the monkeys often made the first saccade towards to the location of the eyes regardless of image content. Our results suggest that the initial fixation placement in faces is driven predominantly by "top–down" or internal factors, such as the prior knowledge of the location of "eyes" within the context of a face.

Keywords Eye movements · Faces · Eyes · Monkey

Introduction

Visual exploration of the world around us involves a series of saccadic eye movements and fixations, and we tend to concentrate our fixations on interesting and informative regions in the scene (Yarbus 1967). The choice of the potential fixation targets can be driven by both bottom–up exogenous or external factors and top–down endogenous or internal factors. External factors are image immanent features, such as local image contrast and local image structure, which transiently attract eye gaze, independent of a particular task. Internal factors, such as an individual's attentional state, expectation, experience and memory, are top–down and task-dependent (Noton and Stark 1971; Mannan et al. 1997; Henderson 2003). It is argued that in general, the initial saccade to an image is driven predominantly by external factors (Parkhurst et al. 2002; Peters et al. 2005), but can also be biased by internal factors (Henderson 2003).

As faces can provide visual information about an individual's gender, age and familiarity, and their expressions provide significant cues to intention and mental state, the ability to recognize these cues and to respond accordingly plays a crucial role in our social communication (Bruce and Young 1998; Emery 2000). Just like humans, rhesus monkeys are sensitive to faces of conspecifics. They are able to discriminate faces of unfamiliar individuals after only a short exposure to sets of their images (Parr et al. 2000). Viewing of faces is accompanied by longer fixations compared with natural scenes (Guo et al. 2006), and is typically associated with a stereotypical eye scanning patterns (Keating and Keating 1982; Nahm et al. 1997; Guo et al. 2003; Gothard et al. 2004; Ghazanfar et al. 2006). Specifically, the eye region in neutral, expressive or vocalizing faces is often the first destination of the saccade and attracts a disproportionate share of fixations compared with other local facial features, suggesting its dominant saliency in the faces.

However, it remains unclear whether this interest in eyes, especially at the earliest stage of face exploration, is

K. Guo (✉)
Research Centre for Comparative Cognition,
Department of Psychology, University of Lincoln,
Lincoln, LN6 7TS, UK
e-mail: kguo@lincoln.ac.uk

221	1038	xxxx	Dispatch: 26.6.07	No. of Pages: 5
Journal	Article	MS Code	LE ☐ TYPESET ☐ CP ☑ DISK ☑	

attributable solely to its local structure properties, or may derive from the knowledge/memory of its location and essence in facial communication. To address this question, in this experiment we systemically manipulated the local image structures of inner face components (i.e. eye region), and compared rhesus monkeys' eye scanning patterns when viewing original monkey face and modified face images. Our results suggested that the top–down guidance (i.e. prior knowledge of the location of the eyes in the faces) plays a crucial role in the saliency of the eye region during early stage of face exploration.

Methods

Subjects

Two male adult rhesus monkeys (*Macaca mulatta*, 4.5–6.0 kg) were used in this study. Initially they were trained to fixate a spot on a computer screen for several seconds in a dimming fixation detection task (Guo et al. 2003). For the purpose of recording eye movements, a scleral eye coil and head restraint were then implanted under aseptic conditions. Throughout the period of the recordings, the animal's weight and general health were monitored daily. All procedures complied with the "Principles of laboratory animal care" (NIH publication no. 86–23, revised 1985) and UK Home Office regulations.

Stimuli and apparatus

Digitized grey scale images were presented through a VSG 2/3 graphics system (Cambridge Research Systems) and displayed on a high frequency non-interlaced gamma-corrected color monitor (110 Hz frame rate, Sony GDM-F500T9) with the resolution of 1,024 × 768 pixels. At a viewing distance of 57 cm the monitor subtended a visual angle of 40° × 30°. The mean luminance of uniform grey background was kept at 6.0 cd/m^2.

Twenty neutral monkey (*Macaca mulatta*) face images were used as stimuli. All images (512 × 512 pixels, 256 grey-levels) were gamma-corrected. For each original face image, we created two scrambled versions (scrambling eye region only, scrambling whole face except for eye region) with the same first- and second-order statistics (image properties determined by the amplitudes of the Fourier spectrum) but different higher-order correlations (image properties determined by the phases of the Fourier spectrum). This was done by computing the Fourier transform over the scrambled facial features and randomizing the phase spectrum (0–2π) in the frequency domain. The Fourier amplitude spectrum of the images was not affected by this procedure. Without higher-order statistical structures corresponding to the sparse distributions of local features, these scrambled image regions lack any visual objects and have a cloud-like appearance (see Fig. 1 for examples), although they have the same mean luminance and root-mean-square contrast as the corresponding facial features (Guo et al. 2005).

In total, three different classes of images were presented to monkeys: (1) 20 original face images, (2) 20 face images with scrambled eyes (eyes scrambled); (3) 20 scrambled faces except for eye regions (eyes only). All images were displayed once in a random order at the center of the screen with a resolution of 512 × 512 pixels (20° × 20°).

During the experiments the monkey was seated in a purpose-built primate chair with head restrained, and viewed the display binocularly. To calibrate eye movement signals, a small red fixation point (FP) (0.2° diameter, 7.8 cd/m^2 luminance) was displayed randomly at 1 of 25 positions (5 × 5 matrix) across the monitor. The distance between adjacent FP positions was 5°. The monkey was trained to follow the FP and maintain fixation for 1 s. After the calibration procedure, the trial was started with a FP displayed on the center of monitor. If the monkey maintained fixation for 500 ms, the FP disappeared and a face image was presented for 10 s. During the presentation, the monkeys passively viewed the images. No reinforcement was given during this procedure, neither were the animals trained on any other task with these stimuli, which could have potentially affected the structure of their behaviour. It was considered that with their lack of training, and in the absence of instrumental responding, their behavior should be as natural as possible.

Fig. 1 Examples of static grey scale monkey face images used in the recording. From *left* to *right*: original face image, face image with scrambled eyes, scrambled face except for eye region

Journal	Article	MS Code	Dispatch: 26.6.07	No. of Pages: 5
221	1038	xxxx	LE ☐ TYPESET ☐	CP ☑ DISK ☑

Eye movement recordings and analysis

Horizontal and vertical eye positions were measured using an 18-in. cubic scleral search coil assembly with 6 min arc sensitivity (CNC Engineering). Eye movement signals were amplified and sampled at 500 Hz through CED1401 plus digital interface (Cambridge Electronic Design). The software developed in Matlab computed horizontal and vertical eye displacement signals as a function of time to determine eye velocity and position. Fixation locations and durations were then extracted from the raw eye tracking data using velocity (less than 0.2° eye displacement at a velocity of less than 20°/s) and duration (greater than 50 ms) criteria (Guo et al. 2006).

Results

Not surprisingly, the original face images were the most salient to the monkeys. They attracted longer viewing time (1-way ANOVA, $F_{(2,129)} = 22.36$, $P = 4.6E-9$) and more fixations (1-way ANOVA, $F_{(2,129)} = 30.01$, $P = 1.98E-11$) than the modified face images (Fig. 2a, b). The two monkeys spent 68 ± 3% (mean ± SEM) of the 10-s image presentation time viewing the original faces, making 17.11 ± 0.9 fixations across the images. The proportion of time spent viewing the image decreased to 44 ± 3 and 39 ± 4%, and the number of fixations declined to 11.25 ± 0.78 and 8.2 ± 0.8 for the eyes scrambled and eyes only images.

Among local facial features, the eye region in original faces often receives the highest proportion of fixations during face exploration (Guo et al. 2003). In this experiment, the eye region in eyes scrambled and eyes only images still attracted a substantial amount of attention, although the cumulative viewing time (Fig. 2c) and the number of fixations (Fig. 2d) were decreased in these two conditions (viewing time: original face 2.91 ± 0.3 s, eyes scrambled 1 ± 0.13 s, eyes only 1.96 ± 0.32 s, 1-way ANOVA, $F_{(2,129)} = 12.81$, $P = 8.44E-6$; number of fixations: original face 8.73 ± 0.65, eyes scrambled 3.59 ± 0.38, eyes only 4.89 ± 0.57, 1-way ANOVA, $F_{(2,129)} = 24.11$, $P = 1.27E-9$). When the same data in Fig. 2c, d was expressed as the percentage of face viewing time (Fig. 2E) and as the proportion of the number of fixations within the images (Fig. 2f), the eye region in original face and eyes only images received the same proportion of face viewing time and fixations (viewing time: original face 41 ± 3%, eyes only 43 ± 4%, Tukey's least significant procedure, $P = 0.6$; fixations: original face 52 ± 3%, eyes only 55 ± 4%, Tukey's least significant procedure, $P = 0.47$). The unrecognisable eyes in eyes scrambled face images attracted less proportion of face viewing time (23 ± 3%, 1-way ANOVA, $F_{(2,129)} = 12.15$, $P = 1.46E-5$) and fixations (32 ± 3%, 1-way ANOVA, $F_{(2,129)} = 14.66$, $P = 1.83E-6$).

To examine whether there were any differences in the spatial distribution of sequential fixation placement during image exploration, we compared the first five fixation placements in each image (this number was chosen as it represented the maximum number of saccades for some images). The probability of fixation placement in the eye region as a function of fixation sequence is plotted in Fig. 3. The eyes had a very higher probability as the first saccade destination (>90%) once the image was presented, even when they were unrecognisable in the eyes scrambled images. For the next four saccades, they had the same

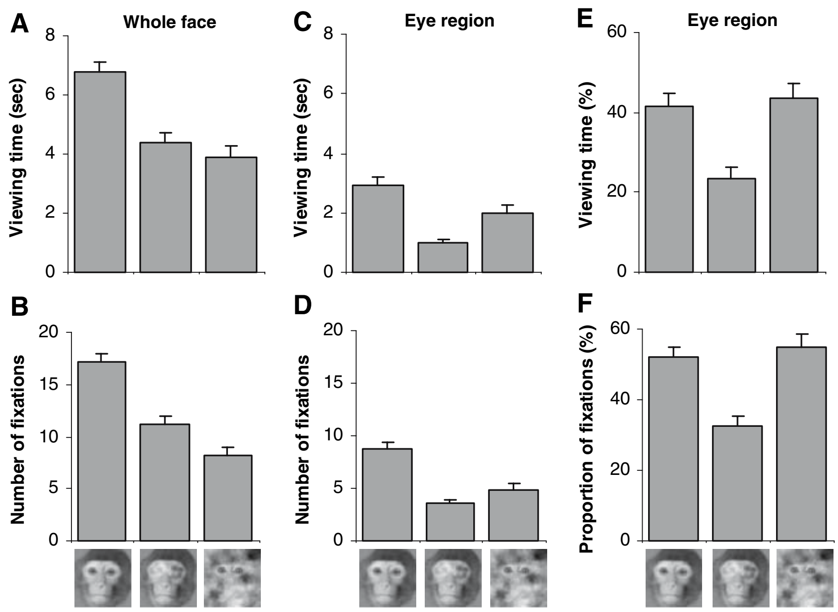

Fig. 2 **a**, **b** Cumulative viewing time and number of fixations within original face, eyes scrambled and eyes only images. **c**, **d** Cumulative viewing time and number of fixations for the eye region within original face, eyes scrambled and eyes only images. **e**, **f** Proportion of cumulative face viewing time and number of fixations for the eye region within original face, eyes scrambled and eyes only images. *Errors bars* indicate standard error of mean

221	1038	XXXX	Dispatch: 26.6.07	No . of Pages: 5
Journal	Article	MS Code	LE ☐ TYPESET ☐	CP ☑ DISK ☑

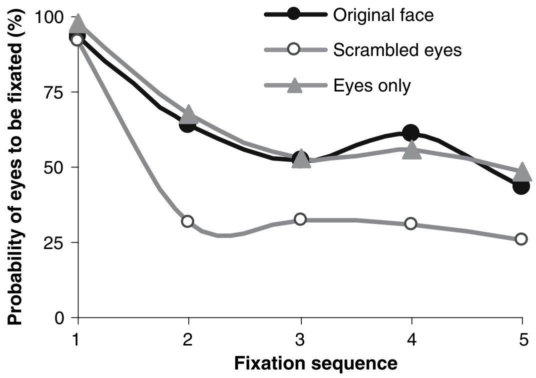

Fig. 3 The probability of the eye region as the destination of first five saccades measured while viewing original face, eyes scrambled and eyes only images

probability to be fixated in the original face and eyes only images (Kolmogorov–Smirnov test, $P > 0.05$), but much less chance to be inspected in the eyes scrambled images (Kolmogorov–Smirnov test, $P < 0.05$).

Discussion

Faces are probably the most important visual stimuli in primate social communications (Bruce and Young 1998). The saliency of the face, however, is dependent on appropriate facial configurations. Disruptions such as inverting faces or randomly rearranging local facial components would reduce the amount of viewing time and fixations directed to the faces (Guo et al. 2003). Here we further observed that the saliency of the face was decreased with the manipulation of local facial structures by scrambling eyes or non-eye regions (Fig. 2a, b), suggesting that the selection of the face as a target for fixation would depend on the prior knowledge concerning the likelihood of the occurrence of the faces (Carpenter and Williams 1995), such as faces presented in a given orientation and within a given context.

Among local facial components, eye region is the most attended feature. During face exploration, both human and non-human primates demonstrated an exaggerated interest in the eye region of the faces of conspecifics (Yarbus 1967; Keating and Keating 1982; Nahm et al. 1997; Guo et al. 2003; Gothard et al. 2004; Ghazanfar et al. 2006). This preferential interest in the eyes remained when the eyes or the rest of the facial structures were scrambled. The eye region in the scrambled faces attracted the same proportion of viewing time and fixations as it did in the original faces, even the unrecognisable eyes in the eyes scrambled images attracted substantial proportion of viewing time (~23%) and fixations (~32%, Fig. 2e, f). Taken together, it seems that both the intrinsic structure (e.g. local contrast or local edges) of the eye region and the knowledge of its location and essence in facial communication contribute to its salience during face exploration. However, the declined cumulative viewing time and fixation numbers towards the eye region in the eyes scrambled and eyes only images (Fig. 2c, d) suggest that the eyes are better processed in concert with other facial features.

The eye region is often the first fixation target following the appearance of the faces (Guo et al. 2003). It is argued that in general, the initial saccade to an image is driven predominantly by "bottom–up" process or external factors such as local image contrast (Parkhurst et al. 2002; Peters et al. 2005). From this perspective, the saliency of the eye region at the earliest stage of face inspection could be attributed to its local structure properties (i.e. the eye region has relatively higher local contrast in grey scale images). However, in our test condition of face images with scrambled eyes, the eye region was also inspected first even when it was unrecognisable (Fig. 3), suggesting that the visual system may retain prior knowledge of the location of "eyes" within the context of a face from past experience, and this knowledge could bias the destination of the initial saccade. In addition, when the face images were inverted or the position of the eyes were rearranged within the faces, the time into the trial for the first saccade directed at the eyes was significantly delayed, indicating the first saccade within the image was not directed at the eyes although their local image properties (contrast and structure) were unaltered (Guo et al. 2003). Taken together, it seems the initial fixation placement in a face image is driven predominantly by "top–down" guidance or internal factors, in particularly the prior knowledge of the location of "eyes" within the context of a normal face. Furthermore, it could be the global semantic characteristics of the faces that determine the initial fixation placement rather than the local semantic characteristics of the eyes.

Our results are consistent with previous behavioural, psychophysical, neurophysiological and neuroimaging studies on the role of the eyes in social interaction in humans and non-human primates. The eyes are one of the first points of contact between infants and mothers, and play a pivotal role in identity recognition and emotional communication (Bruce and Young 1998). They often provide "early warning signals" for rapid assessment and response to salient and potential harmful events (i.e. through the process of joint attention), hence may capture attention involuntarily (Langton et al. 2000; Rauschenberger 2003). While presented alone, the eyes can selectively activate neurons in superior temporal sulcus and amygdala, sometimes with the same response amplitude as the presentation of whole face (Emery 2000; Ghazanfar and Santos 2004). Furthermore, the observation that the eyes do not carry the same relevance for human and monkey infants as human and monkey adults (Thomsen 1974; Farroni et al.

221	1038	xxxx	Dispatch: 26.6.07	No . of Pages: 5
Journal	Article	MS Code	LE ☐ TYPESET ☐	CP ☑ DISK ☑

2002) suggests that the sensitivity to the eyes is a learnt mechanism. Given these considerations, it is reasonable to assume that the knowledge of the location of eye region within a face and its social relevance contribute significantly to its saliency at the earliest stage of face exploration even without specific task demands.

References

Bruce V, Young A (1998) In the eye of the beholder. Oxford University Press, New York

Carpenter RH, Williams ML (1995) Neural computation of log likelihood in control of saccadic eye movements. Nature 377:59–62

Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24:581–604

Farroni T, Csibra G, Simion F, Johnson MH (2002) Eye contact detection in human from birth. Proc Natl Acad Sci USA 99:9602–9605

Ghazanfar AA, Santos LR (2004) Primate brains in the wild: the sensory bases for social interactions. Nat Rev Neurosci 5:603–616

Ghazanfar AA, Nielsen K, Logothetis NK (2006) Eye movements of monkey observers viewing vocalizing conspecifics. Cognition 101:515–529

Gothard KM, Erickson CA, Amaral DG (2004) How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? Anim Cogn 7:25–36

Guo K, Robertson RG, Mahmoodi S, Tadmor Y, Young MP (2003) How do monkeys view faces?—a study of eye movements. Exp Brain Res 150:363–374

Guo K, Robertson RG, Mahmoodi S, Young MP (2005) Centre–surround interactions in response to natural scene stimulation in the primary visual cortex. Eur J Neurosci 21:536–548

Guo K, Mahmoodi S, Robertson RG, Young MP (2006) Longer fixation duration while viewing face images. Exp Brain Res 171:91–98

Henderson JM (2003) Human gaze control during real-world scene perception. Trends Cog Sci 7:498–504

Keating CF, Keating EG (1982) Visual scan patterns of rhesus monkeys viewing faces. Perception 11:211–219

Langton SR, Watt RJ, Bruce II (2000) Do the eyes have it? Cues to the direction of social attention. Trends Cogn Sci 4:50–59

Mannan SK, Ruddock KH, Woodings DS (1997) The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. Spatial Vis 10:165–188

Nahm FKD, Perret A, Amaral DG, Albright TD (1997) How do monkeys look at faces? J Cogn Neurosci 9:611–623

Noton D, Stark L (1971) Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vis Res 11:929–942

Parkhurst D, Law K, Niebur E (2002) Modelling the role of salience in the allocation of overt visual attention. Vis Res 42:107–123

Parr LA, Winslow JT, Hopkins WD (2000) Recognizing facial cues: individual discrimination by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). J Comp Psychol 114:1–14

Peters RJ, Iver A, Itti L, Koch C (2005) Components of bottom–up gaze allocation in natural images. Vision Res 45:2397–2416

Rauschenberger R (2003) Attentional capture by auto- and allo-cues. Psychon Bull Rev 10:814–842

Thomsen CE (1974) Eye contact by non-human primates toward a human observer. Anim Behav 22:144–149

Yarbus A (1967) Eye movements and vision. Plenum, New York

Springer

221	1038	XXXX	Dispatch: 26.6.07	No . of Pages: 5
Journal	Article	MS Code	LE ☐ TYPESET ☐	CP ☑ DISK ☑