CORE
Metadata, citation and similar papers at core.ac.uk
Provided by University of Lincoln Institutional Repository

Longer fixation duration while viewing face images

Kun Guo [1,2], Sasan Mahmoodi [1], Robert G. Robertson [1] and Malcolm P. Young [1]

[1]Department of Psychology, University of Newcastle, Newcastle upon Tyne, NE2 4HH, UK [2]Department of Psychology, University of Lincoln, Lincoln LN6 7TS, UK

Corresponding author:

Dr. Kun Guo

Department of Psychology, University of Lincoln

Brayford Pool, Lincoln, LN6 7TS, UK

Email address: kguo@lincoln.ac.uk

Tel: +44-(0)1522-886294

Fax: +44-(0)1522-886026

Acknowledgements

This work is supported by Wellcome Trust, HFSPO and EU FP5.

Abstract

The spatio-temporal properties of saccadic eye movements can be influenced by the cognitive demand and the characteristics of the observed scene. Probably due to its crucial role in social communication, it is argued that face perception may involve different cognitive processes compared with non-face object or scene perception. In this study, we investigated whether and how face and natural scene images can influence the patterns of visuomotor activity. We recorded monkeys' saccadic eye movements as they freely viewed monkey face and natural scene images. The face and natural scene images attracted similar number of fixations, but viewing of faces was accompanied by longer fixations compared with natural scenes. These longer fixations were dependent on the context of facial features. The duration of fixations directed at facial contours decreased when the face images were scrambled, and increased at the later stage of normal face viewing. The results suggest that face and natural scene images can generate different patterns of visuomotor activity. The extra fixation duration on faces may be correlated with the detailed analysis of facial features.

Keywords: Eye movement – Face image – Natural scene – Monkey

Introduction

Visual exploration of a complex scene involves a series of saccades and fixations, which can shift our attention between specific objects or informative features within the scene and make detailed analysis and identification of the scene (Biederman 1987; Henderson and Hollingworth 1999). There are two important aspects of eye movements while studying gaze control during the scene perception, where fixations tend to be directed (fixation position) and how long they typically remain there (fixation duration) (Henderson 2003). Although human saccadic eye movements show a variety of stereotypic patterns while inspecting visual scenes (Yarbus 1967), the frequency and size of saccades can be modulated by the cognitive demand and characteristics of the observed scene (Salthouse et al. 1981; Jacobs 1986; Pollatsek et al. 1986; Epelboim et al. 1995; Hooge and Erkelens 1998; Andrews and Coppola 1999). For example, longer fixations are normally associated with difficult words in reading task (Pollatsek et al. 1986) and decreased discriminability of target in visual search task (Jacobs 1986; Hooge and Erkelens 1998); and natural scenes generate shorter fixations and larger saccades compared with simple pattern images in free viewing task (Andrews and Coppola 1999).

As faces can provide visual information about an individual's gender, age and familiarity, and their expressions offer significant cues to intention and mental state (Bruce and Young 1998; Emery 2000), the ability to recognize these cues and to respond accordingly plays an important role in the social life of higher primates (Andrew 1963; Anderson 1998). It is argued that face perception is involved in a unique cognitive process compared with non-face object or scene perception. For example, psychophysical studies have observed detrimental recognition performance for inverted faces rather than non-face objects or scenes (face inversion effect) (e.g.

Yin, 1969; Valentine 1988; Rossion and Gauthier 2002), a visual preference for face-like stimuli in human neonates (Johnson and Morton 1991; see also Turati et al. 2002), and selective impairments of face and object recognition in neurological patients (prosopagnosia and visual agnosia) (e.g. Sergent and Signoret 1992; Farah 1996; Moscovitch et al. 1997). Recordings of human event-related potentials showed a different topography to face (including human and animal faces) and non-face object or scene stimuli in the N170 time window (e.g. Bentin et al. 1996; Itier and Taylor 2004; Rousselet et al. 2004). Elecrtophysiology and brain imaging studies further suggested a distinct neuroanatomical region in cerebral cortex associated with the cortical processing of faces (face-selective neurons in monkey inferotemporal cortex, fusiform face area in human cortex) (e.g. Sergent et al. 1992; McCarthy et al. 1997; Tanaka 1997; Tsao et al. 2003). However, this view is recently challenged by some brain imaging studies suggesting that faces are processed by a domain-general system for fine-grained, exemplar-level object perception but probably at different level of recognition or different degree of perceptual expertise (Gauthier et al. 1999, 2000; Tarr and Cheng 2003).

It is not clear, however, whether inspection of face and non-face scenes, which have different image characteristics and may involve different cognitive processes (i.e. different cortical processes, different level of recognition or different degree of perceptual expertise), can influence the patterns of visuomotor activity. To examine this issue, we compared monkeys' saccadic eye movements when they freely viewed face and natural scene images. Familiar scenes sampled from monkeys' daily environment were also used to examine potential influence of the familiarity of natural scene images. This exploratory project is not only important to increase our understanding of the relation between the category of real world stimuli and the

organisation of goal-directed eye movements in non-human primates, but also for comparison with findings from humans, as the behaviour and neurophysiology of monkeys comprises the most significant model for the advancement of research into human brain function. We observed that the face images tended to generate longer fixations compared with the natural scene images, and these longer fixations were associated with the context of facial features.

Methods

Subjects

Three male adult rhesus monkeys (*Macaca mulatta*, 4.5-6.0 kg) were trained to fixate a small fixation point (FP) for several seconds in a dimming fixation detection task. To make eye movement recordings, a scleral eye coil and head restraint were implanted under aseptic conditions (Guo and Benson 1998). 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 gray 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 (6.0 cd/m^2 background luminance, 110 Hz frame rate, Sony GDM-F500T9) with the resolution of 1024 × 768 pixels. At a viewing distance of 57cm the monitor subtended a visual angle of 40 × 30°.

Four different classes of images were used as stimuli (see examples in Fig. 1A and 2A): (1) 20 neutral monkey (*Macaca mulatta*) face images, (2) 20 natural scene images (including buildings, landscape, trees and plants etc.), (3) 15 familiar natural scene images which were taken from monkeys' daily environment, (4) 10 scrambled

monkey face images. The scrambled images were generated by dividing each complete face image into a 4 × 4 matrix and randomly rearranging the parts (Guo et al. 2003). By doing so, most of the local facial features (eyes, nose and mouth) were kept intact and recognizable, but the global structure of the face was disrupted. All images were in sharp focus at all depths of field, and were gamma-corrected and 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 sat in a primate chair with head restrained, and viewed the display binocularly. To calibrate eye movement signals, a small red FP (0.2° diameter, 7.8 cd/m^2 luminance) was displayed randomly at one of twenty-five 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 an image was presented for 20 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 behavior. It was considered that with their lack of training, and in the absence of instrumental responding, their behavior should be as natural as possible.

Eye movement recordings and analysis

Horizontal and vertical eye positions were measured using an 18-inch 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. 2003).

As the main experimental design comprised three levels of image category (faces vs natural scenes vs familiar scenes), one-way repeated analysis of variance (ANOVA) was carried out after pooling the data from three monkeys. Appropriate post hoc testing of differences between levels of image category (Tukey's least significant procedure) was also carried out following detection of significant overall variable ratios.

Results

The gray scale face and natural scene images appeared equally salient to the monkeys. No difference was observed in the number of fixations across the image categories (ANOVA, $F_{(2,162)}$=0.5, p=0.61; Fig. 1B). During the entire 20-second presentation, three monkeys made 24.73 ± 1.51 (Mean ± SEM), 24.82 ± 1.69 and 22.82 ± 1.58 fixations across the face, familiar scene and natural scene images.

The fixation durations were influenced by the image categories. Although frequency distribution analysis showed that the monkeys made frequent short fixations (peak around 200 ms) while viewing the images (Guo et al. 2003), the faces tended to generate longer fixations (ANOVA, $F_{(2,3975)}$=35.7, p=4.29E-16; post-hoc test, face vs familiar scene: p=7.91E-13, face vs natural scene: p=1.71E-11; Fig. 1C). In contrast, the familiar scenes and natural scenes had indistinguishable fixation durations (post-hoc test, p=0.66). The mean fixation durations were 317 ± 8 (Mean ±

SEM), 249 ± 5 and 253 ± 5 ms for face, familiar scene and natural scene images. The conclusion also holds for the median fixation durations which are less sensitive for the skewed distributions of fixation durations (e.g. Fig. 3B in Guo et al., 2003). The median fixation durations were 222, 205 and 200 ms for face, familiar scene and natural scene images.

Inspection of the natural scene is accompanied by a series of fixations directed towards important and informative scene regions. Recent studies observed higher local luminance contrast and lower local two-point correlation for fixated scene patches than un-fixated patches (Reinagel and Zador 1999; Krieger et al. 2000; Parkhurst and Niebur 2003), suggesting that local image statistics, such as luminance contrast, is a major contributor to the saliency map for overt attention (Parkhurst et al. 2002). To examine whether the differences in fixation durations for the three classes of images were due to the differences in the physical properties and statistics of those fixated image regions, we calculated local luminance contrasts around individual fixations in different images. The local contrast is a measure of variability of the intensity within an image patch, and is defined as the standard deviation of the luminance within a square image divided by the mean intensity of the whole image (Reinagel and Zador 1999; Einhäuser and König 2003). The size of the square region was chosen to be 2° × 2° (±1° around the fixation) which roughly covers the spatial scale of the size of the fovea. While the average fixation duration in the face images was longer than that in the familiar scenes (Fig. 1C), the average local contrast around the fixations in the face images (0.2568 ± 0.0034) was not significantly different from that in the familiar scenes (0.2539 ± 0.0038; t-test, $p>0.05$; Fig. 2). However, the average local contrast around the fixations in the natural scene images (0.3512 ± 0.0061) was higher than that in the face and familiar scene images (ANOVA,

$F_{(2,3975)}$=157.11, P=2.63E-66). This is due to the physical properties of the natural scene images, as the average local contrast from random samples in the natural scenes (25 samples per image) was also proportionally higher than that in the face and familiar scene images (ANOVA, $F_{(2,1372)}$=113.02, p=3.67E-46; Fig. 2).

For individual fixations sampled while viewing face, familiar scene and natural scene images, we further plotted its duration against its local contrast (Fig. 3). In agreement with previous study of human subjects (Einhäuser and König 2003), over all images and all subjects, we found no correlation between local contrast and fixation duration (r = 0.00005, 0.0007 and 0.0002 for face, familiar scene and natural scene images). This holds also true for the local contrasts calculated using smaller (1° × 1°) or larger (3° × 3°) spatial scale around the fixations (r<0.001 for all images). This analysis shows that the local luminance contrast was unlikely related to the differences in the fixation durations while viewing face, familiar scene and natural scene images.

As the measurement of local contrast is insensitive to the spatial organization of intensities within an image patch, we also employed two-point correlation function, which calculates the correlation between the point at the centre of each fixation and a point within local neighbourhood of the fixation (±1° around the fixation in this study), to quantify the correlation in intensity between pairs of pixels in the image patch (Reinagel and Zador 1999). The mean and covariance of correlation matrices over the fixations within individual face, familiar scene and natural scene images were calculated and further averaged over each class of the images and subjects (Cootes and Taylor 1992; Cootes et al. 1992). Figure 4 shows the mean of correlations for each class of images. In general, correlation is a function of distance between

image points (pixels). The local image structures around the fixations in the natural scene images seemed to be less correlated than that in the face images.

To further quantify the variations of correlations for each class of images, eigenvalues and eigenvectors of the covariance matrix were computed to analyse principle components of our correlation data over each class (Kreyszig 1999). The Mahalanobis (weighted) distance between the mean of each class and the mean of other classes were finally calculated to determine whether different classes were overlapped with each other or separated from each other (Cootes and Taylor 1992; Cootes et al. 1992). Figure 5 shows the distribution of our data for these three classes of images by considering first two important modes (components) of variations. The distribution function was assumed as a multidimensional Gaussian function whose variances correspond to the eigenvalues of the covariance of the correlation data. These Gaussian functions were considered in a feature space obtained by applying Hotelling transform to our data (Cootes and Taylor 1992; Cootes et al. 1992; Kreyszig 1999). This analysis shows a clear difference in spatial correlations between fixations sampled from the face and natural scene images. The local image structures are more spatially correlated in the face images. However, this difference in local spatial correlations between the face and natural scene images is unlikely related to the difference in fixation durations while viewing the face and natural scene images. Compared with the face images, the correlations between nearby pixels were weak in the natural scene images, indicating a rich structure on small spatial scale in the natural scene images. Therefore the natural scene images are statistically less redundant (Field 1987; Ruderman and Bialek 1994; Simoncelli and Olshausen 2001), and consequently should attract longer fixation durations for the purpose of foveal analysis rather than shorter fixation durations as we observed in the recording.

However, the relationship between fixation duration and local spatial structure of the stimulus may well be task dependent. For example, the natural scene image could attract longer fixation durations in a search task compared with the free viewing task we employed in this experiment. Nevertheless, our observation suggests that the fixation duration is dependent upon not only simple local properties like contrast and spatial correlation, but also some complex features like informativeness.

While viewing the faces, the monkeys' fixation was mainly directed to the principal local facial features, even with the scrambled faces (see examples in Fig. 1A and 6A) (Guo et al. 2003). To investigate whether the longer fixations on facial features are dependent upon their spatial configurations, we compared the durations of fixations on eyes, nose, mouth and facial contours (including hairlines) within normal and scrambled face images (Fig. 6A). While the fixations on eyes, nose and mouth had the same durations between normal and scrambled faces (paired t-test, $p>0.05$), the mean duration of fixations on facial contours of normal faces (302 ± 12 ms) was longer than that of scrambled faces (282 ± 20 ms) (paired t-test, $p=0.03$).

We further compared the durations of each of the first seven fixations on the eyes and facial contours within normal face images (this number was chosen as it represented the maximum number of fixations within the region for some images, Fig. 6B). While the fixation durations on the eyes were the same with changing fixation sequence (ANOVA, $F_{(6,268)}=0.85$, $p=0.53$), the duration of fixations on the facial contours increased gradually at the later stage of fixation (ANOVA, $F_{(6,214)}=3.75$, $p=0.001$). There was no significant change of the fixation durations on the same regions within scrambled faces with increasing fixation sequence (ANOVA, eyes: $F_{(6,98)}=1.25$, $p=0.29$; facial contours: $F_{(6,115)}=0.67$, $p=0.68$).

Discussion

In the present study, we compared the patterns of saccadic eye movements while monkeys freely viewed face and natural scene images (including familiar and novel natural scenes). The face and natural scene images appeared equally salient to the monkeys. They attracted similar number of fixations during the image presentation. However, viewing of the faces was accompanied by longer fixations compared with the natural scenes. This difference in fixation durations across different classes of images is unlikely to be related to the differences in local physical properties and statistics of these images which was demonstrated by the analysis of local luminance contrast (standard deviation of intensity in a fixation patch, Fig. 2, 3) and local two-point correlation function (intensity of the fixated point and nearby points, Fig. 4, 5) across the different classes of images. Comparison between familiar and novel natural scenes showed that these two classes of natural images attracted similar amount of fixation durations (Fig. 1). Because our familiar scenes were 'artificial' man-made scenes sampled from monkeys' daily environment, and novel natural scenes included both 'artificial' scenes (i.e. buildings) and 'natural' scenes (i.e. plants), it is difficult to exclude the potential influence of the 'naturalness' of scenes on fixation duration without further detailed examination with large sample size. However, as our analysis also revealed that the fixation durations sampled from novel 'natural' scenes (253 ± 7 ms) were not significantly different from those sampled from novel 'artificial' scenes (248 ± 11 ms) (t-test, $p=0.61$), it is unlikely that the potential interaction between familiarity and 'naturalness' of the tested scenes could fully account for our observation of difference in fixation durations between face and natural scene images.

Detailed examination of facial configurations further revealed that the longer fixations on facial contours appeared to be dependent upon the arrangement of these contours into a coherent and recognizable object, namely a face. The duration of the fixations on the same facial contours in the scrambled face images were significantly shorter (Fig. 6). These results suggest that face and natural scene images may generate different patterns of visuomotor activity. The extra fixation duration on faces may be correlated with the detailed analysis of facial features.

Oculomotor strategies are closely linked with the cognitive demand (Epelboim et al. 1995), and the fixation duration has been correlated with the amount of information being processed during foveal analysis (Moffit 1980). Longer fixations are usually associated with extra cognitive demand and/or display complexity (Salthouse et al. 1981; Jacobs 1986; Hooge and Erkelens 1998). For example, individual fixation durations are longer during scene memorization than search (Henderson et al. 1999), or when the image at fixation is reduced by contrast or partially obscured by a noise mask (van Diepen 1995).

One of the major differences between face and natural scene images is that faces have inherent social significance. They are behaviorally relevant visual stimuli for primates, which provide essential information about an individual's gender, age, familiarity, intention and mental state (e.g. Bruce and Young 1998; Emery 2000). When viewing a complex scene containing faces, the highest portion of human fixations is directed to the faces (Yarbus 1967). The local facial features, such as eyes, are not just simple geometric patterns or objects. They also contain significant social communicative signals. Like human, monkeys are also heavily reliant on facial signals for social communication. Based on facial cues alone, they are readily able to respond appropriately to the expressions of other individuals (Mendelson et al. 1982),

to recognize and discriminate the faces of familiar and unfamiliar individuals (Rosenfeld and van Hoesen 1979; Parr et al. 2000). Their visual system also appears to be tuned to the informative facial features (Guo et al. 2003). They showed a preferential interest, high density of fixations and longer fixation durations, to the major local facial features while viewing faces. As local image complexity around the fixations is unlikely account for the differences in fixation durations between the face and natural scene images (Fig. 2, 3, 4, 5), the extra duration of fixations for the faces may be correlated with the extra cognitive demand (i.e. "configural process") which involves detailed analysis of local facial features and perceiving relations among the facial features, and therefore maybe important for acquisition and processing of facial cues, such as identity, expression and gaze direction (Maurer et al. 2002). As we only tested neural face images in a free viewing task in this experiment, in the future study it will be interesting to systematically manipulate social relevance over controlled sets of face images and/or cognitive demand, and to investigate the relations among social perception, cognitive demand and patterns of saccadic eye movements.

Interestingly, the disruption of facial configuration (i.e. scrambled faces) seems only affect the number of fixations (Guo et al. 2003) rather than the duration of fixations on the faces (Fig. 6). The fixations on major local facial features, such as eyes, nose and mouth, had the same durations between normal and scrambled faces, only fixations on facial contours slightly decreased for scrambled faces.

Facial cues, such as identity, gender and attractiveness, are sensitive to the manipulation of metric properties of a face (Burton et al. 1993; Perrett et al. 1994; Fellous 1997). Humans can detect variations in facial metric information as small as one minute of visual angle, a value close to the limits of acuity (Haig 1984). Previous studies suggest that this facial metric information is explicitly encoded in the brain

and may be used for face perception and discrimination (Yamane et al. 1988; Young and Yamane 1992). Indeed, the responses of face-selective neurons in anterior inferotemporal cortex of macaques are correlated with dimensions relating the hairline to other facial points, especially the eyes, in face discrimination tasks (Young and Yamane 1992). In our study, the observed monkeys' fixations on facial contours may be correlated with the analysis of the properties of facial dimensions, and this process may require extra fixation time. While in the scrambled faces, due to the lack of information about the facial dimensions, the fixations to these features may just provide less sophisticated local information (i.e. local contrast and local contour).

References

Anderson JR (1998) Social stimuli and social rewards in primate learning and cognition. Behav Process 42: 159-175

Andrew RJ (1963) Evolution of facial expressions. Science 142: 1034-1041

Andrews TJ, Coppola DM (1999) Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. Vision Res 39:2947-2953

Bentin S, Allison T, Puce A, Perez E, McCarthy G (1996) Electrophysiological studies of face perception in humans. J Cogn Neurosci 8: 551-565

Biederman I (1987) Recognition-by-components: a theory of human image understanding. Psychol Rev 94:115-147

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

Burton AM, Bruce V, Dench N (1993) What's the difference between men and women? Evidence from facial measurement. Perception 22:153-176

Cootes TF, Taylor CJ, Cooper DH, Graham J (1992) Training Models of Shapes From Sets of Examples. Proceedings of British Machine Vision Conference: 9-18

Cootes TF, Taylor CJ (1992) Active Shape Models, 'Smart Snakes'. Proceedings of British Machine Vision Conference: 266-275

Einhäuser W, König P (2003) Does luminance-contrast contribute to a saliency map for overt visual attention? Eur J Neurosci 17: 1089-1097

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

Epelboim J, Steinman R, Kowler E, Edwards M, Pizlo Z, Erkelens C, Collewijn H (1995) The function of visual search and memory in sequential looking tasks. Vision Res 35:3401–3422

Farah MJ (1996) Is face recognition 'special'? Evidence from neuropsychology. Behav Brain Res 76:181-189

Fellous JM (1997) Gender discrimination and prediction on the basis of facial metric information. Vision Res 37:1961-1973

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A 4: 2379-2394.

Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci 3: 191-197

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci 2: 568–573

Guo K, Benson PJ (1998) Involuntary eye movements in response to first- and second-order motion. Neuroreport 9:3543-3548

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

Haig ND (1984) The effect of feature displacement on face recognition. Perception 13:505-512.

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

Henderson JM, Hollingworth A (1999) High-level scene perception. Annu Rev Psychol 50:243-271

Henderson JM, Weeks PA, Hollingworth A (1999) Effects of semantic consistency on eye movements during scene viewing. J Exp Psychol Hum Percept Perform 25: 210-228

Hooge IThC, Erkelens CJ (1998) Adjustment of fixation duration in visual search. Vision Res 38:1295–1302

Itier RJ, Taylor MJ (2004) N170 or N1? Spatialtemporal differences between object and face processing using ERPs. Cerebr Cortex 14: 132-142

Jacobs AM (1986) Eye-movement control in visual search: How direct is visual span control? Percept Psychophys 39:47–58

Johnson MH, Morton J (1991) Biology and Cognitive Development: The Case of Face Recognition, Blackwell

Kreyszig E (1999) Advanced Engineering Mathematics. John Wiley, New York

Krieger G, Rentschler I, Hauske G, Schill K, Zetzsche C (2000) Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. Spat Vis 13: 201-214

Maurer D, Le Grand R, Mondloch CJ (2002) The many faces of configural processing. Trends Cognit Sci 6:255-260

McCarthy G, Puce A, Gore JC, Allison T (1997) Face-specific processing in the human fusiform gyrus. J Cognit Neurosci 9:605–610

Mendelson MJ, Haith MM, Goldman-Rakic PS (1982) Face scanning and responsiveness to social cues in infant rhesus monkeys. Dev Psychol 18: 222-228

Moffit K (1980) Evaluation of fixation duration in visual search. Percept Psychophys 274:370-372

Moscovitch M, Winocur G, Behrmann M (1997) What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. J Cognit Neurosci 9: 555-604

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

Parkhurst DJ, Niebur E (2003) Scene content selected by active vision. Spat Vis 16:125-154

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

Perrett DI, May KA, Yoshikawa S (1994) Facial shape and judgements of female attractiveness. Nature 368:239-242

Pollatsek A, Rayner K, Balota DA (1986) Inferences about eye movement control from the perceptual span in reading. Percept Psychophys 40:123–130

Reinagel P, Zador AM (1999) Natural scene statistics at the centre of gaze. Network Comput Neural Syst 10: 341-350

Rosenfeld SA, Van Hoesen GW (1979) Face recognition in the rhesus monkey. Neuropsychologia 17: 503-509

Rossion B, Gauthier I (2002) How does the brain process upright and inverted faces? Behav Cognit Neurosci Rev 1: 63-75

Rousselet GA, Mace MJ, Fabre-Thorpe M (2004) Spatiotemporal analyses of the N170 for human faces, animal faces and objects in natural scenes. Neuroreport 15: 2607-2611

Ruderman DL, Bialek W (1994) Statistics of natural images: scaling in the woods. Phys Rev Lett 73: 814-817

Salthouse TA, Ellis CL, Diener DC, Somberg BL (1981) Stimulus processing during eye fixation. J Exp Psychol Hum Percept Perform 73:611-623

Sergent J, Signoret JL (1992) Varieties of functional deficits in prosopagnosia. Cerebral Cortex 2: 375-388

Sergent J, Otha S, MacDonald B (1992) Functional neuroanatomy of face and object processing. A positron emission tomography study. Brain 115: 15-36

Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24: 1193-1216

Tanaka K (1997) Mechanisms of visual object recognition: monkey and human studies. Curr Opin Neurobio 7: 523-529

Tarr MJ, Cheng YD (2003) Learning to see faces and objects. Trends Cognit Sci 7: 23-30

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. Nat Neurosci 6: 989-995

Turati C, Simion F, Milani I, Umiltà C (2002) Newborns' preference for faces: what is crucial? Dev Psychol 38: 875-882

van Diepen PMJ (1995) Chronometry of foveal information extraction during scene perception. In Findlay JM, Walker R, Kentridge RW (eds) Eye Movement Research: Mechanisms, Processes and Applications. Elsevier, North-Holland, pp 349-362

Valentine T (1988) Upside-down faces: a review of the effects of inversion upon face recognition. Br J Psychol 79: 471-491

Yamane S, Kaji S, Kawano K (1988) What facial features activate face neurons in the inferotemporal cortex of the monkey? Exp Brain Res 73:209-214

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

Yin RK (1969) Looking at upside-down faces. J Exp Psychol 81: 41-45

Young MP, Yamane S (1992) Sparse population coding of faces in inferior temporal cortex. Science 256:1327-1331

Legends

Figure 1. (A) Examples of static gray scale face, familiar scene and natural scene images used in the recording. The white dots within the images indicate the position of each fixation sampled during the image presentation. (B and C) Number of fixations (B) and fixation duration (C) measured while viewing face, familiar scene and natural scene images. Error bars mean standard error of mean.

Figure 2. The average local contrast around the fixations while viewing face, familiar scene and natural scene images (left white columns), and the average local contrast from random samples in face, familiar scene and natural scene images (25 samples per image, left grey columns). Error bars mean standard error of mean.

Figure 3. Dependence of fixation duration on local luminance contrast in face, familiar scene and natural scene images.

Figure 4. The mean of correlations over two-degree image patches around the fixations in face (left), familiar scene (middle) and natural scene images (right). In the far right side of the figure, a scale is presented to indicate the brightness with the corresponding correlation values.

Figure 5. Two dimensional Gaussian functions in feature space corresponding to face, familiar scene and natural scene images. The two axes correspond to the two most

important variations in the covariance matrix, and the units of the axes indicate standard deviations of modes.

Figure 6. (A) Comparison of the durations of fixations on facial contours, eyes, nose and mouth region within normal and scrambled face images. Error bars indicate standard error of the mean. The top graphs are examples of normal and scrambled face images used in the experiment. (B) The change of the fixation durations with increasing fixation sequence at the eyes and facial contours within normal face images.

Figure 1

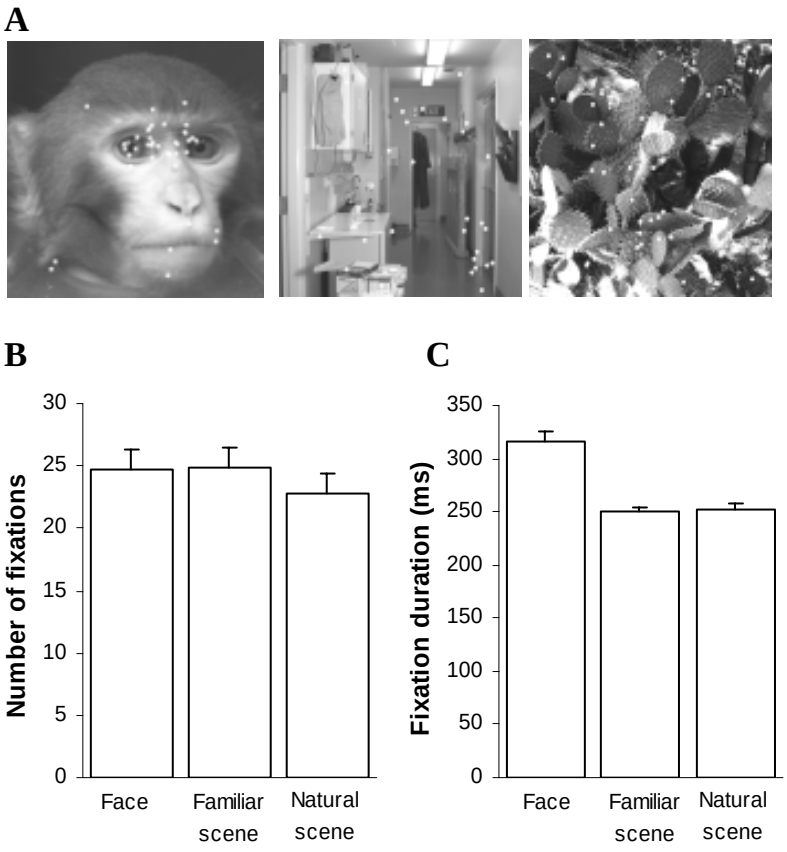

A
B
C
Number of fixations
30
25
20
15
10
5
0
Face
Familiar scene
Natural scene
Fixation duration (ms)
350
300
250
200
150
100
50
0
Face
Familiar scene
Natural scene

Figure 2.

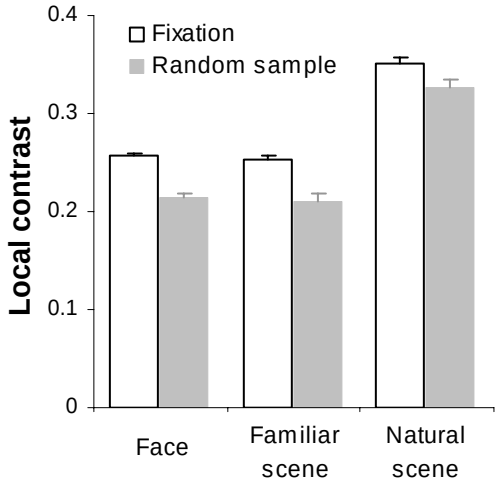

0.4
0.3
0.2
0.1
0
Local contrast
Fixation
Random sample
Face
Familiar scene
Natural scene

Figure 3

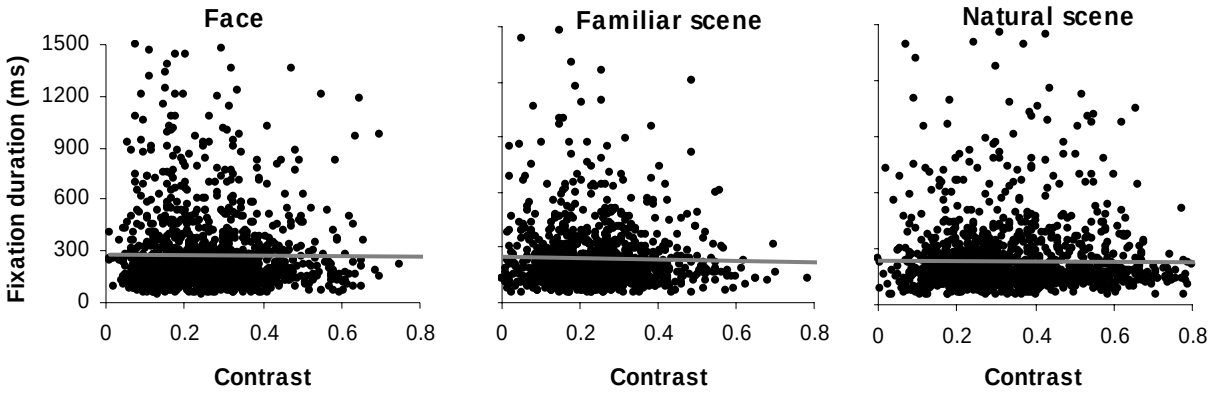

Face
Familiar scene
Natural scene
Fixation duration (ms)
1500
1200
900
600
300
0
0
0.2
0.4
0.6
0.8
Contrast

Figure 4

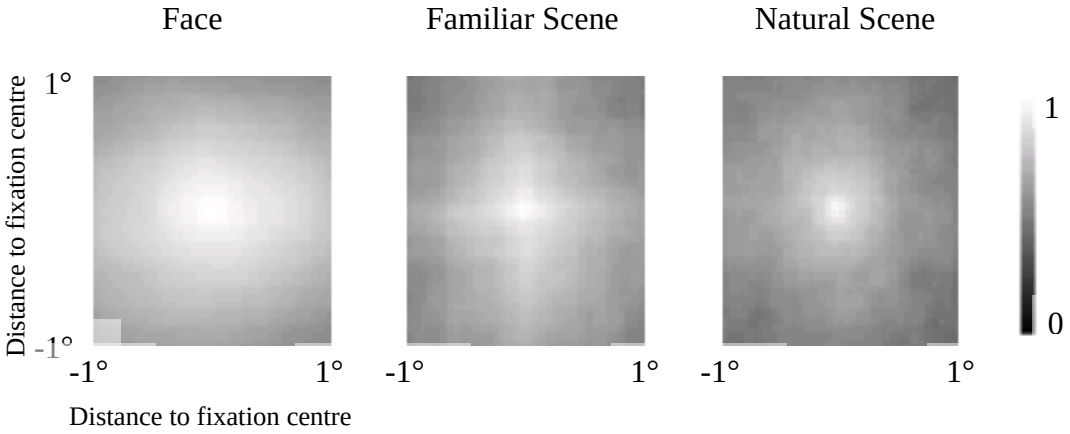

Face
Familiar Scene
Natural Scene
Distance to fixation centre
1°
-1°
-1°
1°
Distance to fixation centre
-1°
1°
-1°
1°
1
0

Figure 5

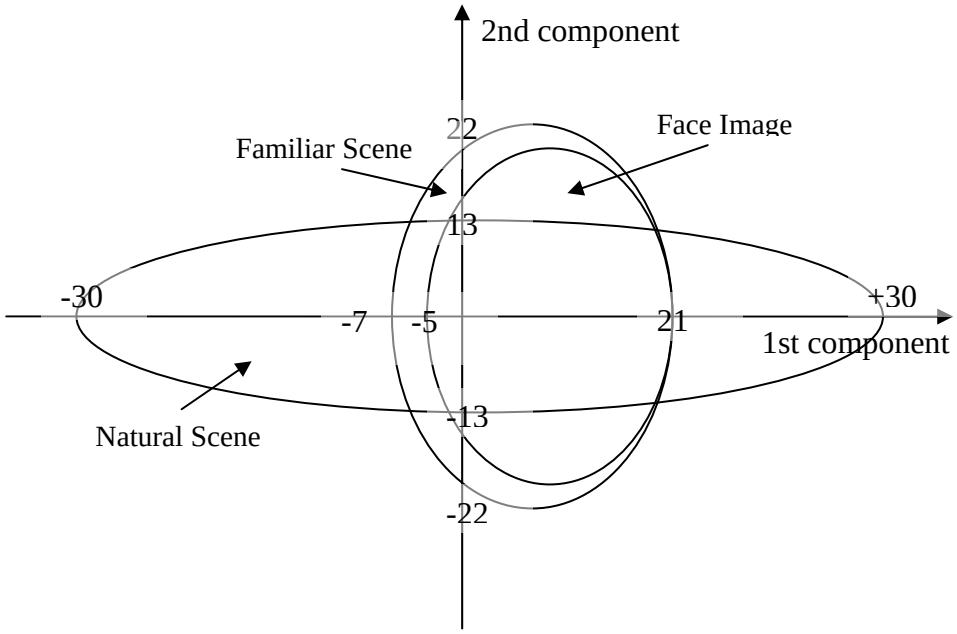
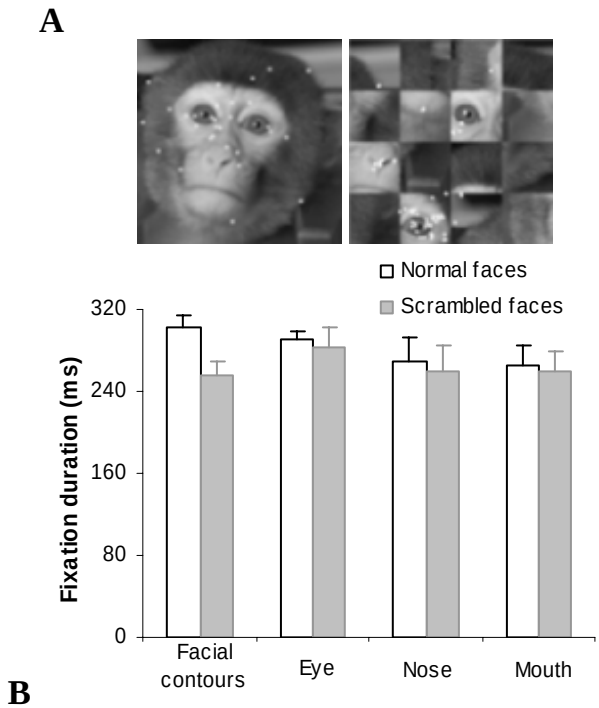

2nd component
22
Face Image
Familiar Scene
13
-30
+30
-7
-5
21
1st component
-13
Natural Scene
-22

Figure 6


A
Normal faces
Scrambled faces
Fixation duration (ms)
320
240
160
80
0
Facial contours
Eye
Nose
Mouth
B
Eyes
Facial contours
Fixation duration (ms)
400
300
200
100
0
0
1
2
3
4
5
6
7
Fixation sequence