The robustness of perception

Authors:
A.A.T.S. Reinders, J.A. den Boer & C. Büchel

(Submitted)
Abstract

The natural environment around us, which is often crowded, cluttered or even foggy, is subject to a dynamically changing composition of objects and events. The human brain is continuously perceiving, recognizing and evaluating this dynamic scene composition. If the perception of degraded visual objects is important, e.g. in case of potential threat stimuli, the brain should be more sensitive in detecting these objects from the natural environment. Therefore, reacting to the dynamically changing environment was hypothesized to involve a robust and quick processing of salient information, which can either be with or without conscious awareness. We investigated the dynamics and robustness of perception using pictures of three salience levels, i.e. fearful faces (most salient), neutral faces (salient) and houses (non-salient), which are appearing from dynamically decreasing random visual noise. Stimuli were matched for luminance, contrast, brightness and spatial frequency information. Reaction times show a significantly earlier response for faces than for houses. Fearful faces were significantly faster detected than neutral faces. The neural correlates sustaining robust perception were investigated with event-related functional magnetic resonance imaging (fMRI). The amygdala showed a significant perception related response for faces, as compared to houses, that was further enhanced for fearful faces, as compared to neutral faces. Our data indicate that emotionally salient information processing is (i) mediated by the amygdala, (ii) more robust than for non-salient stimuli as it shows a significantly lower perceptual threshold.

Introduction

A few studies directly investigated complex natural stimuli by using film clips to explore cerebral responses to a dynamic environment (Karama et al., 2002; Eugene et al., 2003; Levesque et al., 2003; Hasson et al., 2004). However, in a typical functional neuroimaging study, to identify the neural mechanisms involved in environmental perception, the brain is monitored when participants view stripped down and static versions of the environment. When subjects view static pictures or photographs of objects, scenes, letters or faces in a sequence of frames, the visual input has a large degree of perceptual clarity and constancy. Clearly, this does not coincide with a real life scene around us, which is a complex and changing multi-object scene and might often be crowded, cluttered or even foggy.

In highly controlled experimental settings several visual presentation parameters and/or image properties have been manipulated to simulate less optimal perceptual situations of the visual environment. Techniques like backward masking (Morris et al., 1998; Whalen et al., 1998; Grill-Spector et al., 2000; Bar et al.,
Introduction

2001; Sheline et al., 2001), impoverishment by binarizing images (Dolan et al., 1997), image segment scrambling (Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Grill-Spector et al., 1998; Kourtzi and Kanwisher, 2000; Rainer et al., 2002), gradually revealing objects using panels and noise (James et al., 2000), adding grey-level noise to letters and faces (Tarkiainen et al., 2002) or luminance contrast manipulation by changing dot density (Kleinschmidt et al., 2002) have been used. Usually, the (degraded) images are presented in a static and blocked fashion, however, dynamic stimulus presentation has been used to investigate the dynamics of object (letters) recognition (Kleinschmidt et al., 2002), or 3-D object recognition (James et al., 2000).

Despite these studies, the robustness of the visual perception of emotional salient information as compared to emotional non-salient information in a dynamically changing environment is still unknown. Visual information is consciously processed on a cortical level via the striate cortex. However, for the fast detection of emotional relevant stimuli the amygdala is especially suited (LeDoux, 2000). In addition, it has been proposed that coarse salient information is passed via a subcortical pathway to the amygdala (Vuilleumier et al., 2003), without accessing conscious awareness. More specifically, visual information can be passed to the amygdala through the thalamus, bypassing the visual cortex, via the retinal-collicular-pulvinar pathway (de Gelder et al., 1999, 2001). Perceiving fear through this subcortical pathway (Morris et al., 1999; Vuilleumier et al., 2003) allows for a subconscious response through the amygdala. Thus, a response in the amygdala can be triggered by fearful faces even when these faces are not consciously perceived (Morris et al., 1998; Whalen et al., 1998).

We continuously monitored brain activity of human volunteers who were looking at a dynamically changing visual stimulus. This stimulus is an intermediate between a complex moving environmental scene (Hasson et al., 2004) and a conventional visual mapping study using highly controlled static components of visual stimuli. We hypothesized (i) that emotionally salient information processing suffers less from a coarse visual representation, meaning more robust, than non-emotional salient information processing. Therefore, when appearing from visual noise, (ii) faces would be perceived more robustly, i.e. with more noise, than houses, and fearful faces would be perceived with even more noise than neutral faces. Finally, we hypothesized that the amygdala (iii) plays a prominent role for detecting emotionally negative, that is fearful, faces and possibly also for neutral faces (Perrett et al., 1992) and therefore would show activation during the perceptual pop-out for faces, and (iv) shows an even stronger pop-out related activation for fearful faces.
Chapter 7 | The robustness of perception

Materials and methods

Subjects

For this study 17 subjects were recruited, who gave their written informed consent to participate in the study, which was approved by the local ethics committee. One subject was excluded due to medication use and one subject was excluded due to excessive movement during the scanning sessions. The data of the remaining 15 subjects (seven male, eight female) with a mean age of 25.8 years (range 18-36 years) were analyzed.

Stimulus selection

Neutral faces and fearful face images were drawn from the Ekman series of facial affect (Ekman and Friesen, 1976; Ekman, 1982). House pictures were taken from standard north European houses of light color and were adjusted to remove additional distracting information, like trees or fences. All pictures were grey-scale pictures. The images were cropped so that house height and width was approximately the same as face height and width.

As previously described (Whalen et al., 1998), a fearful face signals the presence of danger, but not the source, whereas an angry face represents a direct threat. Consequently, a fearful face is conceptionalized as a contextual stimulus and therefore are used in this paradigm. Early functional imaging studies using normal subjects already found that the processing of negative human facial expressions, e.g. fearful faces, involves specific participation of the amygdala (Morris et al., 1996; Breiter et al., 1996; Gur et al., 2002). The amygdala was also found to be involved in a neural network of fear conditioning (LaBar et al., 1998; Büchel et al., 1998). Furthermore, the function of the human amygdala has been studied with different approaches (see for review: Davis and Whalen (2001)).

Neutral faces and houses were chosen as the control stimulus for the fearful faces. Static visual environmental items presented to subjects often consist of faces (emotional or non-emotional) and/or objects. These different environmental categories selectively activate specific focal regions of the brain (see for recent review: Grill-Spector, 2003). A prominent role of object recognition was attributed to the ventral visual pathway (Grill-Spector, 2003) and ventral occipito-temporal brain areas that show a certain degree of specialization, e.g. the fusiform face area (FFA) for faces (Kanwisher et al., 1997), lateral occipital complex (LOC) for objects (Kourtzi et al., 2003) and parahippocampal place area (PPA) for houses (Epstein and Kanwisher, 1998).
Materials and methods

Stimuli

In previous studies, complex manipulations of the physical characteristics of images have been performed (Fine and Jacobs, 2002), like image contrast manipulation (George et al., 1999; Avidan et al., 2002; Kleinschmidt et al., 2002), frequency filtering (Gold et al. (1999b); Vuilleumier et al. (2003), which includes review on behavioral experiments), applying a constant amplitude spectrum (Grill-Spector et al., 1998) with additional Fourier phase scrambling (Rainer and Miller, 2000; Rainer et al., 2001), morphing the Fourier amplitude spectrum (Parraga et al., 2000) or combinations of Gaussian white noise of various spectral densities (Gold et al., 1999a).

Crucially, we used frequency domain methods (Rainer et al., 2001) to scramble the phase values of the images and to generate the sequence of 80 pictures of increasing image information. Our initial stimulus contained only noise, but gradually a neutral face, a fearful face or a house emerged from the noise (see figure 7.1). This visual noise has been created by randomizing the phase spectrum of the images in Fourier space, which results in images with a degree of noise (giving a cloud-like appearance) related to the ratio of original phase information and randomized phase (see supplemental material). The Fourier method applied incorporates that all the noisy pictures for all conditions have the same amplitude spectrum. More specifically, their amplitude spectrum was approximated by an \( f^{-\alpha} \) function, i.e. a fitted amplitude spectrum (see supplemental material, page 115) in Fourier space using all stimulus pictures, to obtain an averaged amplitude spectrum. This implies that all pictures (houses and faces) were adjusted to have the same brightness, luminance and contrast. In addition, at a certain phase coherence level, the images in all three conditions have the same spatial frequency information (see supplemental material, page 115).

For decreasing noise levels, we merged the noise image with the original image information. More precisely, on the basis of linear interpolation between random and natural phase spectrum we have created a range of stimuli, which start with only visual noise and parametrically increases to an image without visual noise in approximately 40 seconds. The increase of original phase, i.e. decrease of noise, followed an exponential function (fraction of original phase is \( (e^{0.025 \times p} - 1)/(e^{0.025 \times 79} - 1) \), \( p \in [0, 79] \)) to ensure a pop-out after two-thirds of the sequence, i.e. approximately 25 seconds on average. These 80 pictures were dynamically presented to obtain a gradual smooth rebuilding of the original image phase values, i.e., the image appeared gradually from dynamically decreasing random visual noise. Somewhere in the stimulus sequence the subjects became aware of a house or a face, i.e. the perceptual pop-out, which they had to indicate with a button press.
Visual stimulus presentation

Images of the three conditions (neutral faces, fearful faces and houses) all consisted of 640 by 480 pixels which are presented on a screen sized in their natural 4:3 ratio. The actual face or house information was presented within the visual focus of 5 degrees. One run of pictures, from noise to the original image, consisted of 80 steps. Every picture was presented for 490 ms and the original image was therefore rebuild in 39.2 seconds. In this stimulus presentation sequence, the first picture consisted of 100 % noise and contained 0 % original information. The last picture consisted of 0 % noise and 100 % original information. During the presentation of the picture sequence of increasing phase-coherence a dot at the center of the screen, serving as point of fixation, infrequently changed its color,
which had to be acknowledged by the volunteer by a button press. This additional task was introduced to keep attention and motor cortex activation at a constant level throughout each run. Furthermore, volunteers indicated the detection of a stimulus ‘out of the noise’ (pop-out) by an additional button press. The pop-out is defined as the perceptual awareness of the stimulus type, i.e. face or house perception. This task was randomly performed with the right or left hand. Six sessions were obtained and every session contained ten stimulus presentation sequences, which were separated by approximately 15 seconds rest.

**Magnetic resonance imaging**

The neural correlates of robust stimulus detection were investigated using functional magnetic resonance imaging (fMRI). Magnetic resonance scanning was performed on a 1.5 Tesla magnetic resonance imaging (MRI) system (Siemens vision, Erlangen, Germany). Six sessions were obtained with a total of 271 (one subject 270) fMRI scans (25 axial slices, 3 mm thick slices each, 1 mm gap). A gradient echo, i.e. echo-planar imaging (EPI), T2* sensitive sequence was used to acquire these scans (TR 2.1 s, echo-time 40 ms, flip angle 90 degrees, matrix 64 x 64), in descending order. The head was positioned to include the amygdala in a standard head coil using foam pads. To minimize motion artifacts, the head of the subject was fixated by vacuum cushions. The first five volumes in each session were discarded to allow for T1 equilibration effects.

**fMRI data pre-processing**

Statistical parametric mapping (SPM99) was used for spatial transformation (realignment, normalization and smoothing) and statistical analysis of the single subject data (Friston et al., 1995b). SPM2 (www.fil.ion.ucl.ac.uk/spm/spm2.html) was used for the group analysis (random effects model). First, the fMRI time series were realigned to the mean, to correct for intra-subject’s head movement. Hereafter, all the scans were transformed into the standard stereo-taxic Montreal neurological institute (MNI) space (Evans et al., 1993; Friston et al., 1995a) with 7x8x7 non-linear basis functions and heavy regularization during the normalization procedure. As a final step in the pre-processing data was spatially smoothed using an isotropic Gaussian kernel of 11 mm FWHM (Full width at half maximum) to compensate for residual variability in anatomical localization between subjects after spatial normalization and to allow for the application of Gaussian random field theory to address corrected statistical inference. The final voxel size was 3x3x3 mm. A high-pass frequency filter (cut-off 120s) and corrections for auto-correlation between scans was applied to the time-series.
Chapter 7 | The robustness of perception

Statistics

Behavioral data

Due to the fact that we had specific *a priori* hypotheses we used simple one-sided paired *t* tests to perform statistical calculations on the behavioral data. We tested for significant differences in reaction times, due to differences in the pop-out button press, between houses and faces (the general hypothesis) and neutral face versus fearful face (the detailed hypothesis).

fMRI data

The significance of task-related region-specific differences in blood oxygenation level dependent (BOLD) response was assessed using multiple linear regression analyses (Friston et al., 1995b). Within the general linear model (GLM) the parameters were estimated for each subject with ten regressors, reflecting seven conditions (see figure 7.2). The first three regressors resembled the perception of a neutral face, fearful face and a house, i.e. time of pop-out until the end of the trial. The length of this box-car regressor was variable due to the variability of the time-point of pop-out (see figure 7.2, part c)). In addition, the condition dependent pop-out effects were modeled as a simple canonical hemodynamic response function (HRF) and its time derivative (figure 7.2, part b)), yielding six regressors. Note that fMRI data was analyzed with respect to the pop-out, i.e., the data analysis was time-locked on the pop-out. To explain variance which is induced by visual stimulation or button presses, an extra block regressor (convolved with an HRF and starting at the first picture and ending at the last picture) was included in the model (figure 7.2, part d)). Therefore the GLM included ten regressors for each subject and an additional regressor for every session. Both types of faces were compared to houses. Face stimuli were expected to activate the fusiform gyrus, i.e. fusiform face area (FFA) (Kanwisher et al., 1997, 1999), and the amygdala (Perrett et al., 1992; Rolls, 2000). The houses were used as control stimuli, which were expected to activate parahippocampal brain areas, i.e. parahippocampal place area (PPA) (Epstein and Kanwisher, 1998; Epstein et al., 1999).

Average contrast images were calculated for each condition (collapsing across sessions within subjects). These were created by multiplying the contrast vector (condition specific and repeated for each session) with the accompanying regressors from the design matrix and additionally averaged across sessions. These single-subject contrast images were then taken to the second level for a one way ANOVA (within-subjects) group analysis (Friston et al., 1999a,b) in SPM2, which allows for an appropriate non-sphericity correction (Glaser et al., 2002) for a valid inference in the group comparisons.
Figure 7.2: The general linear model (GLM) was estimated with ten regressors (see materials and methods). For every run, the phase coherence increases from 0% original phase to 100% original phase, which is shown in part a) of this figure. During this run the image becomes visible to the subject. At a certain moment the subject perceives the image type, i.e. a face or a house, which is indicated with the pop-out arrow. This event is modeled with the standard canonical hemodynamic response function (HRF), depicted in black, accompanied by its time derivative, depicted in grey (see figure b)). After the pop-out, the subjects consciously perceive the visual stimulus, which is modeled with an HRF of variable duration, plotted in figure c). The regressor in figure part d), is a block regressor (convolved with the HRF), to explain variance which is induced by visual stimulation or button presses.
Chapter 7 | The robustness of perception

*T* statistics for the assessment of significant BOLD responses were performed, by setting specific contrasts to the parameter estimates, for testing condition specific effects on the second level. The resulting statistical parametric maps (SPMs), referring to the probabilistic behavior within the Gaussian random field theory (Worsley, 1994), were investigated for hypothesis based activations. In the case of the amygdala and fusiform gyrus, multiple comparisons correction was performed within the volume of interest (a sphere with a volume of 3054 mm³ (Filipek et al., 1994)). For other apparent activations in the brain, i.e. for which we had no *a priori* regional hypothesis, multiple comparisons correction was performed for the entire brain volume analyzed. Therefore, in all cases the statistical values are reported at *p* < 0.05 corrected. For presentation purposes cluster thresholds were set to include the smallest *a priori* hypothesized area.

Results

Behavioral performance

Reaction times for the perceptual pop-out of the stimuli were calculated from the beginning of the experiment, i.e. from the first picture containing only noise. Time until pop-out, as indicated by a button press, for houses was on average 25.99 ± 0.35 s (mean ± s.e.m.). Faces were perceived on average after 23.71 ± 0.34 s (mean ± s.e.m.). The statistical analysis of the behavioral data revealed a significantly earlier detection of emotionally salient images as compared to emotionally non-salient images (*t*(14) = 14.56, *p* < 0.05).

The faces were further divided into neutral and fearful faces and the time of pop-out was compared. The pop-out for neutral faces was on average after 23.83 ± 0.32 s (mean ± s.e.m.) and the pop-out for fearful faces was on average after 23.59 ± 0.37 s (mean ± s.e.m.). Both neutral and fearful faces were detected significantly earlier than houses (*t*(14) = 13.67, *p* < 0.05 and *t*(14) = 13.46, *p* < 0.05, respectively). More specifically, a comparison of the average pop-out for fearful faces to the average pop-out for neutral faces shows that the fearful face is perceived on average 240 ms earlier than the neutral face (see figure 7.3). Although small, this difference was significant (*t*(14) = 1.92, *p* < 0.05).

fMRI data

For the fMRI data analysis three different blocks and three different events pertinent to our research question were defined (see figure 7.2). The three blocks (fearful faces, neutral faces and houses) were defined from the time of the pop-out until the end of the trial (see figure 7.2 c)) and model the simple full perception of the visual stimulus from the moment of pop-out until the end of the presentation.
Results

Figure 7.3: These figures display the behavioral data. The bar graphs represent the button press in seconds, indicating the perceptual awareness of the image-type, i.e. the pop-out, plus and minus the standard error of the mean. When testing both faces against houses, a significant \( p < 0.05 \) earlier response for faces was found, which is shown in the left figure. In addition, testing fearful faces against neutral faces a significantly \( p < 0.05 \) earlier response for fearful faces was found, which is shown in the right figure.

In addition, the three single events (fearful face, neutral face and house pop-out) were modeled by a canonical hemodynamic response function (HRF) and defined at the moment the volunteer indicated a perceptual pop-out by a button press (see figure 7.2 b)).

Full perception of faces versus houses

Upon testing for a larger BOLD response for houses as compared to faces, using the block regressor, bilateral parahippocampal cortex activation was revealed (see figure 7.4). This activation was accompanied by additional bilateral activation of the dorsal extra-striate visual cortex (see table 7.1). Upon testing for a larger BOLD response for perceiving faces than for perceiving houses, using the block regressor, activation in bilateral fusiform gyrus and bilateral amygdala was identified (see figure 7.4).

Pop-out

In a second step we identified brain areas that displayed differential transient responses at the moment of pop-out. Significant pop-out dependent activation for houses as compared to faces was found in the bilateral parahippocampal cortex (see figure 7.5, top).
Figure 7.4: This figure displays brain activation when fully perceiving a house or a face after pop-out. a) Axial slice (Z = -15 mm) showing bilateral parahippocampal activation, which is related to the perception of houses as compared to the perception of faces (p < 0.05 corrected for multiple comparisons). b) Axial slice (Z = -21 mm) showing bilateral amygdala and bilateral fusiform gyri, which is related to the perception of faces (neutral and fearful) as compared to the perception of houses (p < 0.01 uncorrected for multiple comparisons).
Figure 7.5: Brain activations due to the pop-out event, when detecting a house or a face appearing from decreasing random visual noise. The upper SPM (part a) shows bilateral parahippocampal area activations, on an axial slice (Z = -18 mm, \( p < 0.05 \) corrected for multiple comparisons) which is related to the detection of houses as compared to the detection of faces. In part b) the bilateral amygdala responses, related to the detection of faces stimulus as compared to the detection of houses, are shown in the bottom axial slice (Z = -12 mm, \( p < 0.001 \) uncorrected for multiple comparisons).
Chapter 7 | The robustness of perception

Table 7.1: Summary of the fMRI data results, with coordinates presented in MNI space. \( t \) values are presented under 51.75 degrees of freedom. All \( p \) values are corrected for multiple comparisons. Amygdala and fusiform gyrus activations obtained correction for multiple comparisons for the volume of interest (see methods).

<table>
<thead>
<tr>
<th>Brain area</th>
<th>L/R</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>( t )</th>
<th>( p_{corr} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faces &gt; houses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amygdala</td>
<td>L</td>
<td>-21</td>
<td>-9</td>
<td>-18</td>
<td>3.55</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>27</td>
<td>-3</td>
<td>-21</td>
<td>3.75</td>
<td>0.007</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>L</td>
<td>-45</td>
<td>-48</td>
<td>-24</td>
<td>3.27</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>45</td>
<td>-51</td>
<td>-24</td>
<td>4.88</td>
<td>3.02e-04</td>
</tr>
<tr>
<td>Houses &gt; faces</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parahippocampal area</td>
<td>L</td>
<td>-24</td>
<td>-45</td>
<td>-12</td>
<td>9.02</td>
<td>8.19e-08</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>30</td>
<td>-42</td>
<td>-15</td>
<td>10.02</td>
<td>2.46e-09</td>
</tr>
<tr>
<td>Dorsal extra-striate cortex</td>
<td>L</td>
<td>-33</td>
<td>-90</td>
<td>21</td>
<td>6.84</td>
<td>1.45e-04</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>39</td>
<td>-84</td>
<td>21</td>
<td>7.37</td>
<td>2.54e-05</td>
</tr>
<tr>
<td>Pop-out faces &gt; houses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amygdala</td>
<td>L</td>
<td>-33</td>
<td>-6</td>
<td>-24</td>
<td>4.77</td>
<td>4.26e-04</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>27</td>
<td>-12</td>
<td>-12</td>
<td>4.40</td>
<td>0.001</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>L</td>
<td>-45</td>
<td>-42</td>
<td>-24</td>
<td>2.58</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>45</td>
<td>-51</td>
<td>-21</td>
<td>3.28</td>
<td>0.023</td>
</tr>
<tr>
<td>Pop-out houses &gt; faces</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parahippocampal area</td>
<td>L</td>
<td>-27</td>
<td>-42</td>
<td>-15</td>
<td>7.39</td>
<td>2.42e-05</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>30</td>
<td>-39</td>
<td>-18</td>
<td>8.20</td>
<td>1.54e-06</td>
</tr>
<tr>
<td>Pop-out fearful faces &gt; neutral faces</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amygdala</td>
<td>L</td>
<td>-24</td>
<td>3</td>
<td>-24</td>
<td>3.59</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>12</td>
<td>-6</td>
<td>-27</td>
<td>3.73</td>
<td>0.008</td>
</tr>
</tbody>
</table>

\(^1\) = not significant after multiple comparisons correction, only reported because of contralateral significant activation.

\( p_{corr} = p \) value corrected for multiple comparisons.

L/R = left or right hemisphere; (x,y,z) = MNI coordinates in mm
Figure 7.6: Amygdala activation due to the detection of fearful faces, as compared to the detection of neutral faces. The left amygdala is shown on a coronal slice ($Y = 3$ mm, $p < 0.005$ uncorrected for multiple comparisons).

When testing for areas which show a larger pop-out dependent BOLD response for faces as compared to houses we found strong bilateral amygdala activation (figure 7.5, bottom), with the peak on the right being located slightly more posterior (see table 7.1). These activations were accompanied by weaker bilateral fusiform cortex activation (figure 7.5, bottom). Only the right fusiform activation reached significance, while the left fusiform cortex activation showed a trend (see table 7.1).

As shown in our behavioral data, fearful faces are detected earlier, i.e. when embedded in more noise, than neutral faces. The fMRI data was used to identify the neuronal structures responsible for this effect by comparing pop-out related activity for fearful faces and for neutral faces. Upon testing for higher BOLD response, evoked by the pop-out of fearful faces as compared to neutral faces, bilateral amygdala activation was observed (see table 7.1). Interestingly, there was no fusiform cortex activation at the same statistical threshold. The activation of the right amygdala was located more posterior and its peak activation was included in a larger cluster. The left amygdala activation is depicted in figure 7.6.
Discussion

In this study, we investigated the robustness of perception by continuously monitoring brain activity of human volunteers who were looking at emotionally salient and emotionally non-salient stimuli, which were appearing slowly from visual noise. All stimuli (neutral faces, fearful faces and houses) were matched for luminance, contrast, brightness and spatial frequency information and gradually emerged from dynamically decreasing random visual noise to emulate the dynamics of a noisy environment.

Our behavioral data show that emotionally salient visual stimuli, i.e. faces, are processed more robustly, that is can be detected when embedded in more noise, than emotionally non-salient visual stimuli, i.e. houses. Furthermore, increasing saliency, as is present in faces showing negative emotions like fear, further increased the robustness of detection. The neural correlates sustaining robust perception were explored with event-related functional magnetic resonance imaging. An investigation of transient neuronal signal changes at the pop-out of each stimulus revealed that the amygdala is only activated when encountering an emotional stimulus, i.e. face, and is even more activated when the face has a more salient emotional, i.e. fearful, expression. Previously, it has been found that the amygdala contains face sensitive neurons (Perrett et al., 1992; Rolls, 2000). Our findings strongly suggest that it is the amygdala that serves as a robust detector of emotionally salient stimuli when they are hardly visible.

The moment of recognition, i.e. pop-out, was behaviorally obtained to function as a marker for robust stimulus detection. Brain areas mediating the full perceptual awareness and the pop-out were explored using an fMRI blocked and event-related set-up respectively. The brain activation data showed stimulus dependent activation patterns consistent with previous studies. We found a face specific perception and pop-out related activation in the bilateral fusiform gyrus (Kanwisher et al., 1997, 1999) and in bilateral amygdala. (Perrett et al., 1992; Morris et al., 1996; Breiter et al., 1996; Rolls, 2000). House stimuli activated the parahippocampal cortex (Epstein and Kanwisher, 1998; Epstein et al., 1999). Previous studies have shown that the fusiform gyrus is activated specifically by face perception (Kanwisher et al., 1997; Vuilleumier et al., 2001). The full perceptual awareness of faces, as compared houses, revealed bilateral significant fusiform cortex activation, which was accompanied by weaker bilateral amygdala activation. Interestingly, at the time of the perceptual pop-out of faces, as compared to houses, the amygdala was predominantly activated whereas only weaker fusiform activation was observed (see table 7.1). Although it is hard to argue that an activation is more or less significant than another significant activation, the results suggest that the pop-out of faces is more driven by the amygdala and that the amygdala is therefore specifically involved in the detection of facial stimuli in a dynamically changing environment.
visual environment. In addition, this finding is supported by the findings of Vuilleumier and colleagues (Vuilleumier et al., 2003) in which the fusiform gyrus was involved in fine perceptual shape-analysis while it was less sensitive for a coarse representation of faces, i.e. low-frequency faces, that convey less distinctive facial traits.

Based on our data we propose that human observers are capable of a more robust detection when stimuli are emotionally salient. Recent data (Vuilleumier et al., 2003) indicated that visual inputs to the amygdala might be distinct from those to ventral extra-striate visual cortex. The amygdala and cortical systems possibly operate on different subsets of the visual stimulus in terms of spatial frequency. In that study, the amygdala responses to fearful faces were larger for low-frequency faces than for high frequency faces, indicating that the subcortical pathway provides only coarse inputs to the amygdala. Interestingly, our fMRI and behavioral data supports that the amygdala serves in the robust detection of emotional salient information. It is important to note that our stimuli were matched for luminance, contrast, brightness and spatial frequency information (Rainer et al., 2001). In our stimulus configuration the spatial frequency composition of the stimulus was identical throughout the presentation due to the normalized Fourier amplitude spectrum. When the phase information grows gradually more coherent the spatial structure in the stimulus gradually evolves to its original. Nevertheless, an intermediate image containing X % phase coherence incorporates identical spatial frequency information for each condition, i.e. fearful face, neutral face, and house (see supplemental material). This is crucial because other methods of scrambling also change the spatial frequency information of the image and thus confound noise level and spatial frequency information.

Recognizing threat and danger out of a noisy visual scene is essential for a quick defensive response. On the basis of behavioral data we were able to show that, when subjects view an emotionally salient stimulus coming out of visual noise, they respond significantly quicker than when an emotionally non-salient stimulus appears from the noise. In accordance with current theories highlighting the role of the amygdala in the perception of fear (LeDoux, 1996; LeDoux, 2000), we found a strong activation in bilateral amygdala at the moment of pop-out for faces. Importantly, this activation was stronger for fearful faces than for neutral faces. This finding is in line with hypotheses about a distinguishing perceptual role for the left posterolateral amygdala during the visual processing of facial expressions (Breiter et al., 1996).

A recent study using film scripts (Hasson et al., 2004) has shown that brain activation seems to be linked to observing emotionally salient events in the movie (Pessoa, 2004). In addition, Hasson and colleagues (Hasson et al., 2004) have shown that response properties from previous fMRI studies, which use controlled experimental situations, are valid for more real life situations (Pessoa, 2004). Therefore,
we propose that our findings, using constrained visual stimuli in a dynamical stimulus setting, closely resemble cerebral responses on dynamical natural stimuli.

In conclusion, we explored the neural correlates sustaining robust visual perception using a dynamically changing visual stimulus paradigm and functional magnetic resonance imaging. Increasing saliency of the visual stimuli increased the robustness of detection, which was accompanied by bilateral amygdala activation.

Acknowledgements

The authors are indebted to R. Dolan for comments on the data and J.R. de Jong for help in stimulus development and for help on creating the figures and drafting the manuscript. AATSR was supported by the Graduate school of behavioral and cognitive neurosciences (BCN) in Groningen, the Netherlands (see: www.rug.nl/bcn/). CB is supported by Volkswagenstiftung.
References


Chapter 7 | The robustness of perception


Supplemental material

For chapter 7 entitled:

The robustness of perception

Authors:
A.A.T.S. Reinders, J.A. den Boer & C. Büchel
Image space and Fourier space

A picture of for example an object is built from pixels. In this so called image space these pixels have certain values for the colors red, green and blue (the RGB-values). The picture information can be obtained by reading out the RGB values at some position \((x,y)\). Manipulating the RGB values at position \((x,y)\) changes the information in image space and gives a different image.

An alternative view is to regard the picture as the result of the superposition of a set of sinusoidal waves. Each wave has its own phase, amplitude and frequency characteristics. Multiple waves, all with different frequency, phase and amplitude, can be combined to represent the picture. For color pictures three sets of waves (RGB) are required, while for gray-scale images one set is sufficient.

Fourier theory states that in order to describe any picture with width \(w\) and height \(h\) using waves, \(w \times h\) waves with different frequency are required. Both spaces have the same amount of degrees of freedom: \(w \times h\) pixels in image space and \(w \times h\) frequencies in Fourier (wave) space. Since the picture is two dimensional, the waves have two frequency components. A matrix, e.g. \(H\), with the same dimensions as the original picture contains all the required information to describe the picture in Fourier space. Each element of the matrix \(H\) will be a complex number belonging to one of the \(w \times h\) waves. Let \((i, j)\) denote the position of the element in the matrix.

1. Amplitude: \(|H(i, j)|\)
2. Phase: \(\text{arg}(H)\)
3. Frequency: \(f_x = i, f_y = j, f = \sqrt{f_x^2 + f_y^2}\)

In order to construct the Fourier matrix, a Fourier transformation can be utilized. An inverse Fourier transformation translates the Fourier matrix back into the original image matrix, i.e. the matrix containing the gray values of the pixels. In the case of color pictures all three components can be treated separately.

Phase spectrum

Images of the same size contain exactly the same frequencies in Fourier space. However, the amplitude and phase belonging to a certain frequency are in general different between pictures. The spatial structure information (in image space) is specified by the phase belonging to the frequencies (the phase spectrum). The phase spectrum of the Fourier transform of a picture of a house could be replaced by that of a picture of a face. After inverse Fourier transformation, the picture will represent a face. The amplitude spectrum merely influences image brightness, contrast and sharpness.
With respect to spatial frequency \( (f(x,y)) \), natural images have a characteristic \( f^{-\alpha} \) Fourier amplitude spectrum (van der Schaaf and van Hateren, 1996; Rainer et al., 2001), i.e., most of the information is contained in the low frequencies components. Images differ from each other primarily by their phase spectrum and not in terms of the amplitude spectrum. In (natural) images, the phase spectra are highly correlated as a result of spatial structures (edges, corners, etc.). Images with a random phase spectrum have a cloud-like appearance (see figure 7.7). When randomizing a phase spectrum, i.e. scrambling, the specific spatial structures (edges, corners, etc.) are not present, which can provide us with a specific form of visual noise. When interpolating the random phase spectrum with the original phase spectrum, i.e. increase in phase coherence from 0 % to 100 %, the phase returns to its original values. This also means that while increasing the phase coherence the correlation in the phase spectra increases as a result of the increase of spatial structures (edges, corners, etc.). At a certain moment, e.g. at a phase coherence of \( X \) %, a certain amount of spatial structures in the image has returned. The term ‘spatial frequency information’ is defined as the amount of spatial structure in the image.

**Stimulus development**

For all the pictures (houses and faces) the background color was set to one identical grey-value. As a first step we estimated \( \alpha \) from all images used in the experiment, i.e. faces and houses. Natural images have an characteristic \( f^{-\alpha} \) Fourier amplitude spectrum (van der Schaaf and van Hateren, 1996; Rainer et al., 2001). The amplitude spectrum of natural images can be approximated by:

\[
A = C \cdot f^{-\alpha} \tag{7.1}
\]

in which \( A \) is the Fourier amplitude spectrum, \( C \) is a scale factor and \( f \) the frequency. The frequency \( f = 0 \) is an exception, for which \( A \) has the value of the DC component. For each of our 30 pictures (ten houses, 10 neutral faces and ten fearful faces) we obtained the value of \( \alpha \) and DC and consecutively averaged these values to get the average spectrum. More precisely, \( \alpha \) of a picture can be obtained by a least-square-fit through the log-log plot of frequency and amplitude. This procedure was performed in Fourier space. The value of our averaged \( \alpha \) was 1.49. This value is consistent with the value (\( \alpha = 1.4 \)) for natural images (Rainer et al., 2001).

The DC-component is defined as the value in the amplitude spectrum belonging to the zero frequency \((0,0)\), i.e. a constant term. Thus, the DC-component is the ‘brightness’ of the picture (a higher value of the DC gives a brighter image) and is related to the average grey-value/intensity for pixels in the picture, i.e. in image space. Therefore, the DC component can also be seen as the (averaged of all grey-
value in the image)*(size of the image). The scale factor \( C \) (in \( A = C*f^{-\alpha} \)), indicates the contrast of an image. A higher value for \( C \) gives more contrast.

The second step was to obtain an averaged amplitude spectrum on the basis of this averaged \( \alpha \). Using the specific mirror symmetric characteristic of an amplitude spectrum, the \( \alpha \) amplitude spectrum was rebuilt using only the first quadrant. The other three quadrants are mirrored versions of the first. Only even sized pictures (640×480) were used, therefore zero-frequencies (first row and first column) were added specifically and are mirror symmetric. The averaged (of our 30 pictures) DC-component (position \( f(x,y) = (0,0) \)) was used to set the brightness. On basis of visual inspection the scaling factor \( C \), one contrast value for all pictures, was manually set for naturally looking pictures (see figure 7.8). This averaged amplitude spectrum was used for all stimulus pictures.

**Figure 7.7:** One sequence of the stimuli as presented during the fMRI investigation. The upper left corner depicts the original picture. In the lower right corner the rebuild picture with the averaged amplitude spectrum is shown.
Figure 7.8: All pictures (houses and faces) have the same fitted amplitude spectrum, as obtained in Fourier space. This results in identical luminance and contrast for all the pictures (see also figure 7.9), as presented in the fMRI investigation. In the top row of the figure six original images are shown (not used during the fMRI investigation). In the bottom row the same six images are shown, but now with an averaged, i.e. fitted, amplitude spectrum. All the 30 images were used to obtain this averaged amplitude spectrum. Two neutral faces, two fearful faces and two houses are depicted.

The third step was to create the scrambled pictures using the $\alpha$ based averaged amplitude spectrum, as created in step two. In Fourier space, the phase of the pictures is scrambled. This is done in two steps: creating a random phase spectrum and combining this with the original natural phase spectrum. The first step was to create a matrix with random (phase) values. An upper half matrix of uniformly distributed angles between $-\pi : \pi$ was created. Using the point-mirrored-anti-symmetric phase-spectrum characteristics, in combination with one single mirror anti-symmetric row and one column of random values, a full Fourier random phase spectrum was assembled.

Picture rebuilding

After this procedure the original natural phase spectrum and a randomized phase spectrum were available, which were combined using linear interpolation. The increase in information was not linear but was an exponential grading, to avoid too quick detection of the original image. The exponential increase was 0.025 over the range of 80 pictures:

$$\text{fraction of original phase} = \frac{(e^{0.025\cdot p} - 1)}{(e^{0.025\cdot 79} - 1)} \quad p \in [0, 79]$$

(7.2)

The value of 0.025 for the exponential grading was determined with pilot experiments. Using this value for the exponential the pictures became visible around 2/3 of the sequence presentation.
To eliminate artifacts in the corners (resulting in white corner-tips) the original 640×480 sized images, were padded with a grey boundary to size 800×600 before the Fourier transform. After inverse Fourier transformation the image was restored to the original size. Values of the inverse Fourier transformation outside the standard grey-value range were clipped to fit in this [0, 255] range. In other words, values below 0 were put to 0, while values exceeding 255 were set to 255.

The Fourier method applied ensures that all the 80 pictures of the 30 images (ten houses, ten neutral faces and ten fearful faces) have equal luminance and contrast. In addition at a certain phase coherence level, the three conditions have equal spatial frequency information. The controlled physical stimuli characteristics are depicted in figure 7.9. Controlling these physical items ensures that the responses of the subjects are not contaminated by contrast dependent responses (Tiippana and Nasanen, 1999; George et al., 1999; Avidan et al., 2002) or spatial frequency bandwidth dependent responses (Gold et al., 1999; Nasanen, 1999; Vuilleumier et al., 2003).

**Visual stimulus presentation**

Images of the three conditions (neutral face, fearful face and houses) were presented in their natural 4:3 ratio. The actual face or house information was presented within the visual focus of 5 degrees. One run of pictures, from noise to the original image, consisted of 80 steps. Every picture was presented for 490 ms and was therefore rebuilt in 39.2 seconds. In this stimulus presentation sequence, the first picture consisted of 100 % noise and contained 0 % original information. The last picture consisted of 0% noise and 100 % original information (see figure 7.7).
Figure 7.9: The Fourier method applied incorporates that all the 80 pictures of the 30 images (ten houses, ten neutral faces and ten fearful faces) have obtained equal luminance and contrast (see figure 7.8). In addition at a certain phase coherence level, the three conditions have equal spatial frequency information. The controlled physical stimuli characteristics are depicted in this figure. The stimulus presentation sequences of 80 pictures of increasing image information is depicted in the top part of the figure. The increase in phase coherence is indicated by the black bars. We interpolated the scrambled image with the original image information to gradually turn back from random phase spectrum to the original phase spectrum of the image.
Chapter 7 | The robustness of perception

References


