Face Recognition and Cortical Responses Show Similar Sensitivity to Noise Spatial Frequency

To find cortical correlates of face recognition, we manipulated the recognizability of face images in a parametric manner by masking them with narrow-band spatial noise. Face recognition performance was best at the lowest and highest noise spatial frequencies (NSFs, 2 and 45 c/image, respectively), and degraded gradually toward central NSFs (11–16 c/image). The strength of the 130–180 ms neuromagnetic response (M170) in the temporo-occipital cortex paralleled the recognition performance, whereas the mid-occipital response at 70–120 ms acted in the opposite manner, being strongest for the central NSFs. To noise stimuli without faces, M170 was small and rather insensitive to NSF, whereas the mid-occipital responses resembled closely the responses to the combined face and noise stimuli. These results suggest that the 100 ms mid-occipital response is sensitive to the central spatial frequencies that are critical for face recognition, whereas the M170 response is sensitive to the visibility of a face and closely related to face recognition.

Keywords: cortex, face, human, MEG, recognition, visual

Introduction

Several areas of the human cerebral cortex are critical for processing of faces, and they appear to contribute to extraction and analysis of invariant (identity) and variant (expression, eye gaze) facial features (e.g. Haxby et al., 2000). In subjects viewing faces, positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) activations have been observed bilaterally in temporo-occipital areas, especially in the lateral fusiform gyrus (Sergent et al., 1992; Haxby et al., 1994; Clark et al., 1996; Kanwisher et al., 1997; McCarthy et al., 1997), the lateral inferior occipital gyri and the posterior superior temporal sulcus (Kanwisher et al., 1997; Halgren et al., 1999; Haxby et al., 1999). From these activations, those in the lateral fusiform gyrus appear to be associated with perception of face identity (Sergent et al., 1992; George et al., 1999; Hoffman and Haxby, 2000).

In electroencephalographic (EEG) and magnetoencephalographic (MEG) recordings, a response that is at least twice as strong for faces than for any control stimuli tested so far (including textures and a large variety of objects) peaks 140–180 ms after the stimulus onset. Originally this EEG response was reported as the electric vertex-positive peak VPP (Jeffreys, 1989), but later studies have focused on the temporo-occipital surface-negative deflection N170 and its magnetic counterpart N170m or M170 (Lu et al., 1991; Bentin et al., 1996; George et al., 1996; Sams et al., 1997; Halgren et al., 2000; Liu et al., 2000). Source modeling of MEG signals suggests that M170 is generated in the occipito-temporal cortex in the region of the fusiform gyrus (Sams et al., 1997; Halgren et al., 2000); this interpretation agrees with intracranial recordings of the 200 ms face-selective response (Allison et al., 1999). Thus face-specific brain areas can be probed with MEG, but there are no parametric comparisons, yet, between the cortical response strengths and behavioral face recognition performance. For example, Tarkiainen et al. (2002) observed that the M170 amplitude gradually decreased as a function of increased noise amplitude in face images; however, behavioral face recognition or detection performance was not reported. Instead of manipulating the stimuli, Liu et al. (2002) compared the M170 amplitudes during trials associated with successful and unsuccessful face recognition, and showed that M170 was stronger during the successful trials.

Face recognition depends on a limited range of spatial frequencies (SFs), as is evident from studies that applied low-pass, high-pass and band-pass filtering of the images (Fiorentini et al., 1983; Hayes et al., 1986; Peli et al., 1994; Costen et al., 1996), or used masking with plaids (i.e. the sum of a vertical and horizontal grating; Tieger and Ganz, 1979). According to these studies, SFs of ~10–20 cycles per face width (c/face) appear the most important for recognition of face identity. Näsänen (1999) recently applied narrow-band additive spatial noise and observed maximum sensitivity for face recognition at 8–13 c/face, with a bandwidth of slightly less than 2 octaves.

In search for brain processes closely related to face recognition, we looked for cortical signals that would parallel face recognition performance when specific SF bands of the face images are manipulated. We thus recorded MEG responses to images that contained narrow-band spatial noise at 10 different SF bands, with central frequencies from 2 to 45 cycles per image width (c/image; corresponding to 1.5–33 c/face); the subject’s task was to report the occurrences of a target face. The noise amplitude was the same for all noise spatial frequencies (NSFs) and was selected so that face recognition was difficult at NSFs of 11–16 c/image but easy at the lowest and highest NSFs. For comparison, we also recorded responses to noiseless face images with high and low contrast, as well as to noise without faces. Preliminary results have been reported in abstract form (Tanskanen et al., 2002).

Materials and Methods

Subjects

After receiving written informed consent, we studied six healthy members of laboratory personnel (two females and four males; mean age 29 years, range 22–46; five right-handed and one ambidextrous; normal or corrected-to-normal visual acuity). The experimental protocol was accepted by the Ethics Committee of the Helsinki and Uusimaa Hospital District.
Stimuli and Procedure

Equipment

The stimulus images were generated, before the experiments, with custom-made software. Stimulus presentation was controlled by Presentation® software (www.neurobs.com) run on a PC computer. The stimuli were displayed on a rear projection screen (Dataplex 735-DPS0) by a data projector (VistaPro™, Christie Digital Systems Inc., Cypress, CA). The projector is based on Digital Light Processing™ and hosts three digital micromirror panels; thus the luminance onsets and offsets are symmetric and abrupt, and all three colors are drawn simultaneously (for details on the projector performance, see Pack et al., 2001). The high luminance output of the projector was attenuated with a 1.4 log unit neutral-density filter placed in front of the lens, resulting in average luminance of 131 cd/m². The non-linearity of the luminance response of the projector was known and taken into account during stimulus image generation by using its inverse function (gamma correction). The experiments were run in standard VGA mode (resolution 640 x 480 pixels, frame rate 60 Hz, 256 gray levels).

Stimuli

The stimuli were combinations of synthetic facial images and of spatial noise masks with 10 different noise SFs (NSFs). The noise amplitude was always the same, resulting in identical signal-to-noise ratios for different NSFs (see below). We preferred to add noise rather than to filter or phase-randomize the images in selected bands, because strong effects in recognizability of high contrast images can with the latter methods be obtained only after manipulation of very broad SF bands.

A set of eight synthetic face images was adopted from Näsänen (1999). One of the face images was the mean of four real faces, and the other images were obtained by warping this face according to the locations of corresponding points in seven other real photographs in which the poses were highly similar and the facial expressions were neutral; for details, see Näsänen (1999). This procedure kept the texture, lighting, pose and expression relatively constant, and the stimuli differed mainly in shape.

The stimulus images were 256 x 256 pixels in size, which corresponded to 11 x 11 cm² on the screen and to 7 x 7 deg on the viewing distance of 88 cm.

Figure 8 (continuous line) shows the mean amplitude spectrum of these eight face images in log-log scale. The decrease of the amplitude (A) as a function of spatial frequency (f) obeys function A(f) = k f^{-1.9} (dashed line), where k is a constant.

To produce noise masks, narrow-band noise was first obtained by filtering white Gaussian noise with rectangular band-pass Fast-Fourier-Transform filters so that the noise was white within each band with zero power elsewhere. The center SF was either 2, 2.8, 4, 5.6, 8, 11, 16, 23, 32 or 45 c/image, corresponding to 0.28-6.3 c/deg during stimulus presentation, and the bandwidth of the noise was always 2 c/image (0.28 c/deg). For each noise center SF, 20 different noise masks were generated. The contrast of noise was constant at all SF bands.

A noisy face was a weighted sum of a face image and a noise mask. Figure 2 (top row) shows some examples of the 1600 such stimuli (8 faces x 20 masks x 10 NSFs). In addition, ‘low-contrast’ and ‘high-contrast’ sets of eight noiseless faces were presented.

Stimulus contrast was expressed in RMS contrast (c_rms), which was defined as

$$c_{\text{rms}} = \sqrt{\sum_{x} \sum_{y} c^2(x, y)}$$

where A is the area of the image and $c(x, y)$ is the contrast signal, defined as

$c(x, y) = \frac{I(x, y) - I_0}{I_0}$

where $I(x, y)$ is the luminance waveform and $I_0$ is the mean luminance of the image (Legge et al., 1987). The RMS contrast was 0.079 for faces presented with noise and low-contrast noiseless faces, and 0.26 for high-contrast noiseless faces. The corresponding Michelson contrasts were 28 and 93%, respectively. The RMS contrast of noise was 0.16.

For all noisy faces, the signal-to-noise ratio (i.e. the ratio of the RMS contrasts of faces and noise) was 0.5, independently on the noise spatial frequency. This signal-to-noise ratio was selected so that face recognition was difficult at NSFs centered on the critical band for face recognition (11-16 c/image), without too much interference at low and high NSFs (Näsänen, 1999).

Procedure

The subjects viewed the stimuli binocularly in a room illuminated solely by the light from the screen. The display area was gray (80 cd/m²), with a lighter (131 cd/m²) 11 x 11 cm² stimulus presentation area in the middle, equal in mean luminance to the stimuli. The subjects were asked to keep their gaze fixated to the middle of this square.

The stimuli were presented once every 2.5 s, each for a duration of 0.5 s, with abrupt onsets and offsets. All 12 stimulus categories (10 categories with noise masks, plus the high- and low-contrast noiseless faces) were presented within the same blocks in random order. Subjects were asked to respond with a right index finger lift to images representing the target person, indicated before the MEG recording. To avoid movement-related contamination of the data, target trials (12.5% of all) were not included in the MEG analysis.

Before the MEG recordings, the subjects went through a two-phase behavioral training to learn to recognize the target face among the other faces. In the first phase, two noiseless faces were shown next to each other, and the subject had to indicate which of them was the target. Then a stimulus sequence similar to that in the actual experiment was presented, with the exceptions that the probability of the target face was doubled and that feedback was provided after each trial. After a 30-40 min training, all subjects were able to recognize the target person with close to 100% accuracy.

MEG Recording

Whole-scalp neuromagnetic signals were measured in a magnetically shielded room, while the subject was sitting with the head surrounded by the helmet-shaped Vectorview™ 306-channel neuromagnetometer (Neuromag Ltd, Helsinki, Finland). The detector array comprises 102 identical triple sensor units, each housing two planar first-order SQUID (Superconducting Quantum Interference Device) gradiometers and one magnetometer. The two gradiometers of each unit measure orthogonal tangential derivatives of the magnetic field component normal to the head surface. Planar gradiometers pick up the strongest signals just above a locally activated brain area and thereby these regions can be readily used as first guesses of the activated brain areas.

MEG signals were bandpass filtered at 0.1-173 Hz and sampled at 600 Hz. Signals were averaged online over a time interval starting 0.3 s before and ending 1 s after the onset of the stimulus. Vertical and horizontal electro-oculograms were recorded for online rejection of epochs contaminated by blinks and eye movements. A total of 60-90 responses for each of the 12 categories were collected in two or three measurement blocks, 15-20 min each. Thus, each subject saw only 50-75% of the generated stimulus images.

Before the MEG recordings, four head position marker coils were attached to the subject’s scalp. The positions of the coils and of three anatomical landmarks (nasion and points immediately anterior to the ear canals) were measured with a 3-D digitizer (iSpace Fastrak™, Polhemus Inc., Colchester, VT). At the beginning of each recording block, the position of the subject’s head with respect to the sensor array was determined by feeding current to the marker coils. This information was afterwards used for combining the sources of the measured neuromagnetic signals with the subjects’ structural MRIs by first identifying the anatomical landmarks in the MR images.

MEG Data Analysis

The effects of environmental noise on the averaged signals were first attenuated by projecting out noise sub-spaces on the basis of room noise measured in the absence of the subject (Parkkonen et al., 1999). The responses were then digitally low-pass filtered at 35 Hz, and a 300 ms pre-stimulus baseline was applied for amplitude measurements. Only signals from the 204 gradiometers were analyzed.

The averaged evoked responses of each subject were first screened for experimental effects. The responses that showed clear dependence...
on NSF were then modeled with equivalent current dipoles, assuming a spherical volume conductor that was fitted to the posterior part of the intracranial volume (for a detailed description of the method, see Hämmäläinen et al., 1993). These current dipoles served two aims: first, they acted as spatial filters to collapse data of a set of sensors to yield a better signal-to-noise ratio, and second, they gave an idea about the sites of cortical areas where the observed effects took place. The dipole locations and orientations were found by a least-squares fit to a subset of sensors around the local signal maxima.

The dipoles found in the conditions with the strongest signals were then inserted into a multidipole model that was used to reveal source strengths as a function of time in all conditions. It is important to note that the relative source strengths obtained in the multidipole model did not change when the source strengths were extracted separately for single dipole models; therefore we can exclude any harmful interactions within the source model. We also ascertained that the observed NSF dependencies were present for sources identified in conditions that elicited less-than-maximum responses.

The source coordinates were transformed into standard brain coordinates. This alignment was based on a 12 parameter affine transformation of the individual brains (Woods et al., 1998), followed by a refinement with a non-linear elastic transformation (Schormann et al., 1996) to match a standard atlas brain (Roland and Zilles, 1994). As a result, major sulci and other important brain structures were well aligned.

Results

Averaged Cortical Responses

Figure 1 shows averaged neuromagnetic responses of Subject 1 to facial images containing low and medium SF noise. A complex pattern of responses can be seen on sensors located above the occipital, temporal and parietal lobes. Systematic effects of NSFs were seen in the mid-occipital and temporo-occipital regions. The mid-occipital response (A), peaking around 100 ms, was minimal for the lowest sf (red trace), but high for the center NSF (blue trace). On the contrary, the temporo-occipital response (B), peaking around 140 ms, was strongest to images with low NSF. The response disappeared for the medium noise SF; i.e. the condition where face recognition was difficult or impossible.

Main Effects of Noise Spatial Frequency

Figure 2 shows the main effects of NSF on face recognition and cortical responses. The top row illustrates face images with noise masks with center NSFs from 2.0 to 45 c/image. The bars indicate the average face recognition performance of the six subjects for each NSF, measured as the percentage of detected target faces during the MEG recordings. Face recognition was close to perfect for stimuli with the lowest and highest NSFs but poor (0–27% for individual subjects) for those with the middle NSFs. As the noise amplitudes and signal-to-noise ratios were constant across all images, these data demonstrate the reliance of human face perception on SFs of ~11–16 c/image.

Note that the high NSFs are not properly reproduced in the images in Figure 2, which, in addition to the small size of the shown images, makes the most difficult conditions to appear at somewhat lower NSFs than suggested by the recognition data. The original stimuli are available as supplementary material.

The sources of the mid-occipital response at 100 ms and the temporo-occipital response at 140 ms, both showing systematic effects of NSF, were modeled with equivalent current dipoles. The two lowest rows of Figure 2 show the resulting source strengths for a single subject (S1). The mid-occipital source is weakest at the lowest NSFs, then increases as the NSF increases, and is again weaker at the highest NSFs. The source in the left temporo-occipital area shows a nearly opposite behavior, being strong at high and low NSFs but weak at the central NSFs, thereby paralleling the face recognition performance.

Mid-occipital Responses at 100 ms

The mid-occipital 100 ms response was observed in all six subjects, and it was adequately modeled with a current dipole in

**Figure 1.** Averaged responses of Subject 1 to facial images with two noise SFs. Each pair of traces represents responses from the two orthogonal planar gradiometers in a sensor unit (cf. Hämmäläinen et al., 1993). (A) A typical mid-occipital response; (B) a typical temporo-occipital response.

---

528 Face Recognition and Cortical Responses · Tanskanen et al.
and 113 six subjects. The corresponding mean low-contrast noiseless images. From 2.1 to 2.7 octaves (mean 2.5 and the half-amplitude bandwidth of the tuning curve varied response was obtained at NSFs of 11--32 c/image (1.5--4.5 c/deg), in two subjects, two dipoles with different orientations at nearby locations were required to properly account for field variance; however, as the strengths of the two sources showed a very similar dependence on NSF, the dipole with higher signal-to-noise ratio was selected for further analysis.

The mid-occipital response was consistently present down to NSFs of 5.6 c/image. The peak latency of the response was shortest (mean ± SEM = 85 ± 4 ms) at 5.6 c/image and it then systematically prolonged to 86 ± 3, 89 ± 3, 92 ± 2, 96 ± 3, 105 ± 2 and 113 ± 2 ms for NSFs of 8, 11, 16, 23, 32, and 45 c/image, respectively.

Figure 3 (top) shows the mean normalized source strengths for the mid-occipital responses. The smallest responses were elicited by the images with the lowest NSF and by the low-contrast noiseless faces. Around the NSF of 5.6 c/image, the responses started to increase and they reached the maximum on average at 20.5 ± 3.7 c/image (2.9 ± 0.52 c/deg). In different individuals (see Fig. 4, top), the maximum mid-occipital response was obtained at NSFs of 11--32 c/image (1.5--4.5 c/deg), and the half-amplitude bandwidth of the tuning curve varied from 2.1 to 2.7 octaves (mean 2.5 ± 0.1).

The responses then decreased again for the highest NSFs, but with considerable interindividual variability, as can be seen from the large error bars and from the individual data in Figure 4 (top). The responses were statistically significantly (*t* = 5.80, *P* < 0.005, paired two-tailed *t* test) stronger to high-contrast than low-contrast noiseless images.

The MRI insert in Figure 4 shows the source locations for the six subjects. The corresponding mean ± SEM Talairach coordinates (Talairach and Tournoux, 1988) were *x* = 4 ± 3, *y* = -86 ± 4, *z* = -7 ± 6.

**Temporo-occipital Responses at 130–180 ms**

Prominent responses peaked at 130–180 ms, with sources in the temporo-occipital or posterior temporal cortex. In three subjects the NSF effects were stronger in the left than the right hemisphere, and in two subjects they were stronger in the right hemisphere. In one subject, the left and right response strengths were almost identical, and the slightly stronger right-hemisphere source was selected for further analysis.

In all subjects, the latencies of the temporo-occipital responses were shortest for the high-contrast noiseless image, on average 144 ± 5 ms. Responses to the low-contrast noiseless images peaked 10 ± 1 ms later; the corresponding delays were 14 ± 4 ms for responses to stimuli with the highest and lowest NSFs (2, 2.8 and 45 c/image), and 21 ± 5 ms for the 4 and 32 c/image stimuli. When compared with responses to high-contrast images, all these delays were statistically significant (*P* < 0.02).

Figure 3 (middle) shows the mean peak amplitudes for each NSF. The images with low NSF elicited strong signals. At NSF ≥ 4 c/image the amplitudes started to decrease, and were the smallest at NSFs of 8–16 c/image. The signals then increased again for the highest NSFs. Individual data are shown in Figure 4 (bottom).

The responses were smallest at NSFs of 11.9 ± 2.7 c/image (1.7 ± 0.38 c/deg), with a range of 5.6–23 c/image (0.8–3.2 c/deg) across subjects. The half-amplitude bandwidth was 2.3 ± 0.13 octaves, ranging from 2.0 to 2.7 octaves.

The strengths of the mid-occipital and temporo-occipital responses seemed to be inversely related to each other. In line with this, the two subjects (S1 and S6) who had the strongest mid-occipital responses at higher NSFs than the other subjects also had their smallest temporo-occipital responses at higher NSFs than the others. However, in all six subjects, the largest mid-occipital response occurred at a higher NSF than the smallest temporo-occipital response (*P* < 0.03; binomial test).

The MRI insert (Fig. 4, bottom) shows the source locations for the six subjects; the mean ± SEM Talairach coordinates of the temporo-occipital sources were *x* = ±39 ± 2, *y* = -68 ± 4 and *z* = -9 ± 2, and these sources were 33–66 mm (mean 47 mm) apart from the sources of the mid-occipital responses.

**Behavioral Face Recognition versus Cortical Responses**

Figure 3 (bottom) plots the mean ± SEM percentage of target faces detected during the MEG recordings. Stimulus recognition varied as a function of NSF as expected, being close to perfect at NSFs of 2.0 and 45 c/image, as well as for both sets of the
noiseless images. Between these extreme NSFs, the performance declined gradually, and was worst for the faces with NSFs of 11.0–16.0 c/image.

Figure 5 presents the behavioral data (continuous line and diamonds) together with the strengths for the mid-occipital (dotted line, circles) and temporo-occipital (dashed line, squares) sources. The main shapes of the face recognition and temporo-occipital source strength curves resemble each other.

Figure 6 shows that the amplitudes of both the mid-occipital responses (left) and the temporo-occipital responses (right) correlated statistically significantly with the recognition performance ($r = -0.87; P < 0.001$ and $r = 0.89; P < 0.001$, respectively). Correlation between the mid-occipital and temporo-occipital responses was $-0.79 (P < 0.005)$.

**Responses to Plain Noise with Different NSFs**

To further clarify the functional roles of the mid-occipital M100 and the temporo-occipital M170 responses, we measured responses to the noise masks alone in two subjects. Plain noise stimuli and the original face + noise stimuli were presented within the same blocks in random order. To limit the duration of the experiment, only every second NSF used in the main experiment was applied. The subjects’ task and the target probability were identical to the main experiment. Figure 7 (top) shows that the mid-occipital M100 was very similar to plain noise (dashed line) and face + noise (continuous line) stimuli; this result is in line with the very small 100 ms responses elicited by plain faces. On the contrary, the M170 responses (bottom) were strongly affected by the presence of a face. For the face + noise stimuli, M170 shows the same U-shaped modulation as was observed in the main experiment, but for
plain noise, the response is small and almost independent of NSF. If anything, the lowest NSFs might have elicited somewhat stronger activity in the temporo-occipital areas around 130–180 ms than did the central or high NSFs.

Discussion

Our aim was to search for brain correlates of face recognition when the face images were masked with narrow-band noise covering spatial frequencies from 2 to 45 c/image. Two cortical responses showed distinct dependence on the NSF. First, the early mid-occipital responses at 70–120 ms (M100) were smallest for low NSFs, increased until 20 c/image (2.9 c/deg), and decreased again for the highest NSFs. Second, the temporo-

Source Area of the Mid-occipital Response M100

According to our source modeling, the mid-occipital responses are, in all except one subject, generated in the lingual gyrus. This source location, combined with the early peak latency and the strong dependence of the response on spatial frequency, suggests generation of M100 in the retinotopically organized visual cortex. Accordingly, the mean Talairach coordinates of the source area (4, -86, -7) agree with location of the human V1/V2 cortex observed in PET studies (Hasnain et al., 1998).

Source Area of the Temporo-occipital Response M170

The temporo-occipital source, with mean Talairach coordinates of ±39, -68, -9, is within 0.5 cm from the face-selective lateral occipital activity reported in fMRI (43, -65, -4: Puce et al., 1996). The location agrees with the site of the lateral occipital complex (LOC), a region associated with the perception of object shape in fMRI studies (Grill-Spector et al., 1999). Our source is within 10–20 mm from the area in lateral fusiform
cortex that typically shows face-specific activation in fMRI studies (Puce et al., 1996; Kanwisher et al., 1997; McCarthy et al., 1997; Haxby et al., 1999). However, in intracranial EEG recordings, the sites of face-selective 200 ms responses in the ventral temporo-occipital cortex extend for 7 cm in the anterior-posterior and for 4 cm in the mediolateral directions, and simultaneous face-specific sites can be observed on the lateral surface of the temporal lobe (Allison et al., 1999). Although the face-selective fMRI activation and the M170 response may not reflect identical neuronal processes (Furey et al., 2001), the good agreement between fMRI activation sites and our M170 source area suggests major contribution for the M170 response from the lateral temporo-occipital and/or fusiform areas.

**Comparison of Face Recognition, Temporo-occipital Activity and Stimulus Properties**

The present results demonstrate that behavioral face recognition and the cortical M170 face response are sensitive to NSF in a band-pass manner at similar frequencies (Fig. 5): face recognition is most difficult at NSFs of 11–16 c/image and the center critical frequency for M170 is 11.9 ± 2.7 c/image.

These data agree with previous psychophysical findings that critical bands for face recognition range from 10 to 20 c/face (Tieger and Ganz, 1979; Fiorentini et al., 1983; Hayes et al., 1986; Peli et al., 1994; Costen et al., 1996). Our stimuli resembled those by Näätänen (1999) whose two subjects were most sensitive to noise at 11 c/face, corresponding to 15 c/image in the present study.

The observed dependence of face recognition and cortical responses on certain SFs cannot be explained by the properties of the face stimuli only. As the spectrum of the face images (Fig. 8) shows, the lowest SFs of the stimuli had the highest intensity, and the amplitude decreased as a function of spatial frequency. However, both face recognition and the M170 amplitudes were insensitive to noise in the low SFs.

**Relationship between Mid-occipital and Temporo-occipital Responses**

The mid-occipital M100 responses were inversely related to recognition performance, being strong when performance was poor, and vice versa. A similar relation between mid-occipital and temporo-occipital responses was observed by Tarkiainen et al. (2002), whose 100 ms mid-occipital responses increased linearly as a function of the amplitude of pixel noise, whereas the M170 amplitude gradually decreased when the noise amplitude in the face images increased.

Because face recognition is most sensitive to frequencies at which the mid-occipital regions respond strongly, a possible explanation for the observed tuning would be the signal-to-noise ratio of the representation of the face + noise images in V1/V2 cortex, which in turn could affect the output from these regions to the higher-order areas that are important in face and object recognition. In other words, our results would fit with the idea that the NSF sensitivities of face recognition and of face-selective areas reflect properties of the earlier visual areas. However, the center absolute spatial frequency (c/deg) of the critical band for face recognition increases with decreasing stimulus size but slower than in inverse proportion. For example, a fourfold increase in viewing distance and thus in absolute spatial frequencies results in a decrease of the relative spatial frequency (c/face) critical for face recognition from 11 to 8 c/face width (Näätänen, 1999). Thus, the critical band is neither scale invariant nor fixed in absolute spatial frequency. Similarly, the spatial frequency used for letter recognition (expressed in cycles per letter width) decreases with decreasing letter size (Majaj et al., 2002). Furthermore, if the critical band was defined solely by the sensitivity of the early areas, it should be the same for all object categories, an issue that has not been explored yet.

**Band-pass Characteristics of the Mid-occipital Response**

Band-pass characteristics of the mid-occipital 100 ms responses have previously been reported in several studies. Musselwhite and Jeffreys (1985) observed that occipital evoked potentials to black-and-white gratings peak at ~4 c/deg. Fylan et al. (1997) demonstrated that MEG responses to chromatic gratings show band-pass characteristics, with the strongest responses at 1–2 c/deg. The low-frequency attenuation in the mid-occipital responses most likely reflects the receptive field sizes of V1/V2 neurons, which are inversely proportional to the optimal spatial frequency of each neuron; consequently, fewer neurons are needed to cover the stimulus area for low than high spatial frequencies. Indeed, the low-frequency attenuation in the MEG responses can be compensated for by increasing the stimulus area (Fylan et al., 1997). On the other hand, the attenuation of the cortical responses (and behavioral contrast sensitivity) to high spatial frequencies is due to a number of optical and neural factors (for a review, see De Valois and De Valois, 1990).

**Face and Object Recognition and Cortical Responses**

Our result that face recognition correlates well with the M170 amplitude agrees with an earlier report on stronger M170 during trials associated with successful than unsuccessful face recognition (Liu et al., 2002). In the same study, M100 was stronger during successful than unsuccessful face detection (the subjects had to tell whether the stimulus was a face or another object). These findings, similarly as ours, point towards different functional roles of the M100 and M170 responses. One possibility is that M100 reflects an early step in the processing of spatial frequencies that carry critical information about faces.

In the present study, we did not explicitly test for the specificity of M100 and M170 to faces as opposed to other object categories. However, a number of previous studies have
demonstrated that the N170/M170 responses are significantly stronger to faces than to other object categories or textures (Lu et al., 1991; Bentin et al., 1996; George et al., 1996; Sams et al., 1997; Halgren et al., 2000; Liu et al., 2000). As comparisons such as face versus house or face versus checkerboard unavoidably involve steep changes at many levels of the stimuli, we preferred to manipulate the visibility of faces on a continuous scale, and to bind recognition performance and neural responses on this continuum. The present result on the strong positive correlation between M170 and the visibility of a face embedded in noise conforms with the previous reports on face specificity of M170. Instead, M100 was not specific to faces, as it was considerably weaker to plain faces than to plain noise at central spatial frequencies.

According to a previous fMRI study, the strengths of activations in LOC and in the fusiform gyrus anterior to the face-specific region correlate with the successfulness of object recognition (Bar et al., 2001). This interesting parallel to the present findings should, however, be interpreted with caution, because M170 and the hemodynamic response in the fusiform gyrus may reflect different neural processes (Furey et al., 2001). Presumably, processing at the ‘M170 stage’ is a prerequisite for the later, more sustained neural activity reflected by fMRI. More specifically, the processes underlying M170 could reflect perceptual analysis preceding the retrieval of stored information about face identity. Such an interpretation is supported by EEG recordings that using familiar versus unknown or learned versus new faces indicated that familiarity influences only the later, 300–600 ms responses, not the 170 ms deflection (Eimer, 2000; Henson et al., 2003; Paller et al., 2003).

Conclusions

We have demonstrated that both the recognizability of human faces and the amplitudes of cortical evoked responses can be systematically manipulated by adding noise to the images. Face recognition was easy with low and high NSFs but difficult with central NSFs. Two cortical responses showed a distinct dependence on NSF. The 70–120 ms mid-occipital response M100, presumably reflecting the sensitivity of the V1/V2 areas to the NSF, was small at NSFs at which face recognition was easy, and vice versa. On the contrary, the 130–180 ms response (M170) in the temporo-occipital cortex showed a strong correlation with the recognition performance, being largest at the lowest and highest NSFs, and smallest at the medium NSFs. A control experiment confirmed that the mid-occipital response was modulated by the spatial frequency of the noise mask, whereas the temporo-occipital response depended on the visibility of a face. These findings support the importance of neuronal processes underlying the M170 for face recognition.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oupjournals.org/.

Notes

We thank M. Seppä for MRI normalizations and brain coordinate transformations and J.V. Haxby, H. Renvall and S. Vanni for comments on the manuscript. The MRI scans were obtained at the Department of Radiology, Helsinki University Central Hospital. Supported by the Academy of Finland and EU’s Large-Scale Facility Neuro-BIRCH III at Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology.

Address correspondence to T. Tanskanen, Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, PO Box 2200, FIN-02015, Espoo, Finland. Email: topi.tanskanen@hut.fi.

References


