

Dynamics of visual feature analysis and object-level processing in face versus letter-string perception

A. Tarkiainen,¹ P. L. Cornelissen² and R. Salmelin¹

¹Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, Finland and

²Psychology Department, Newcastle University, Newcastle upon Tyne, UK

Correspondence to: Antti Tarkiainen, Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, P.O. Box 2200, Fin-02015 HUT, Finland
E-mail: antti.tarkiainen@hut.fi

Summary

Neurons in the human inferior occipitotemporal cortex respond to specific categories of images, such as numbers, letters and faces, within 150–200 ms. Here we identify the locus in time when stimulus-specific analysis emerges by comparing the dynamics of face and letter-string perception in the same 10 individuals. An ideal paradigm was provided by our previous study on letter-strings, in which noise-masking of stimuli revealed putative visual feature processing at 100 ms around the occipital midline followed by letter-string-specific activation at 150 ms in the left inferior occipitotemporal cortex. In the present study, noise-masking of cartoon-like faces revealed that the response at 100 ms

increased linearly with the visual complexity of the images, a result that was similar for faces and letter-strings. By 150 ms, faces and letter-strings had entered their own stimulus-specific processing routes in the inferior occipitotemporal cortex, with identical timing and large spatial overlap. However, letter-string analysis lateralized to the left hemisphere, whereas face processing occurred more bilaterally or with right-hemisphere preponderance. The inferior occipitotemporal activations at ~150 ms, which take place after the visual feature analysis at ~100 ms, are likely to represent a general object-level analysis stage that acts as a rapid gateway to higher cognitive processing.

Keywords: MEG; extrastriate; facial expression; occipitotemporal; noise masking

Abbreviations: LH = left hemisphere; MEG = magnetoencephalography; RH = right hemisphere

Introduction

Recognition of faces and facial expressions is important for successful communication and social behaviour. Not surprisingly, certain cortical areas are thought to deal specifically with face processing. Haemodynamic functional imaging methods have shown that faces activate certain parts of the occipitotemporal cortex, typically in the fusiform gyrus, more than other types of image (e.g. Sergent *et al.*, 1992; Haxby *et al.*, 1994; Clark *et al.*, 1996; Puce *et al.*, 1996; Kanwisher *et al.*, 1997; McCarthy *et al.*, 1997; Gorno-Tempini *et al.*, 1998). Electromagnetic functional recordings have timed this activation to ~150–200 ms after image onset (e.g. Lu *et al.*, 1991; Allison *et al.*, 1994, 1999; Nobre *et al.*, 1994; Sams *et al.*, 1997; Swithenby *et al.*, 1998; McCarthy *et al.*, 1999; Halgren *et al.*, 2000). Occipitotemporal face-specific activation can be seen bilaterally, but generally with at least slight right-hemisphere (RH) dominance (e.g. Sergent *et al.*, 1992; Haxby *et al.*, 1994; Bentin *et al.*, 1996; Puce *et al.*, 1996; Kanwisher *et al.*, 1997; Sams *et al.*, 1997; Gorno-

Tempini *et al.*, 1998; Swithenby *et al.*, 1998; Halgren *et al.*, 2000). The importance of occipitotemporal face activations is confirmed by observations showing that lesions in the basal occipitotemporal cortex are associated with the inability to recognize familiar faces (prosopagnosia) (for reviews, see Damasio *et al.*, 1990; De Renzi *et al.*, 1994).

Neuronal populations in the inferior occipitotemporal cortex have also been shown to respond preferentially to other behaviourally relevant image categories, such as letter-strings and numbers (Allison *et al.*, 1994), in a time window very similar to that for face processing (Allison *et al.*, 1994; Nobre *et al.*, 1994). However, while the timing seems to be comparable for faces and letter-strings, letter-string processing is rather strongly lateralized to the left hemisphere (LH) (Puce *et al.*, 1996; Salmelin *et al.*, 1996; Kuriki *et al.*, 1998; Tarkiainen *et al.*, 1999), unlike face processing. The similarity of the face- and letter-string-specific occipitotemporal brain activations in timing and location (Allison *et al.*, 1994;

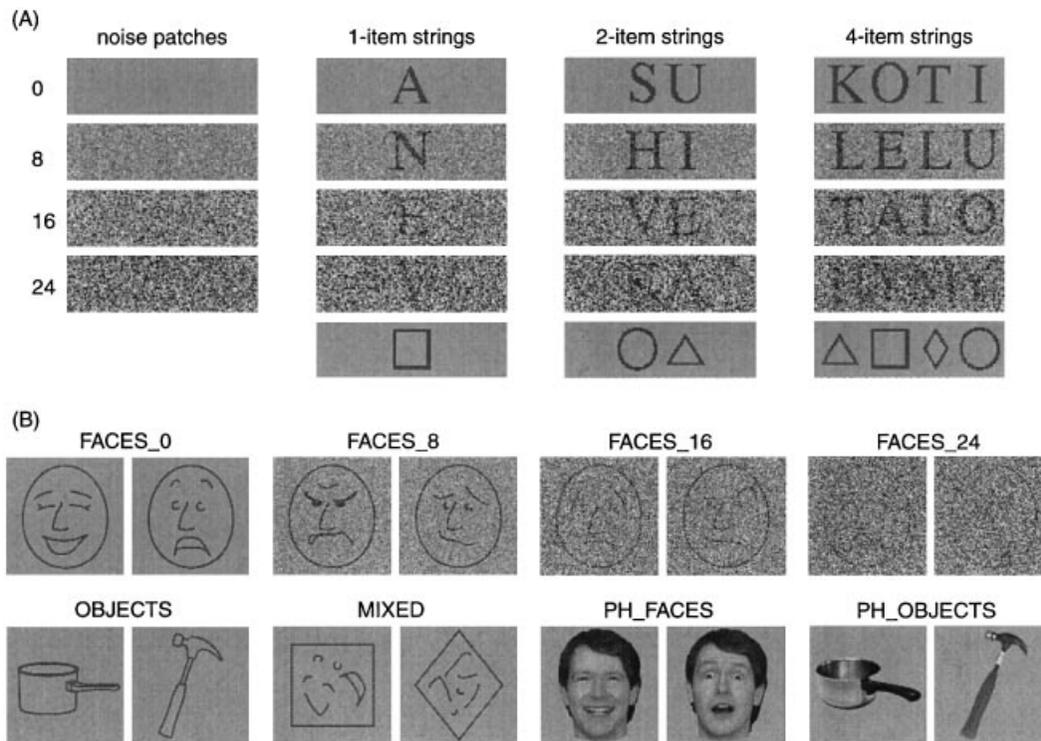


Fig. 1 Examples of stimuli used in the earlier study on letter-string reading (A) and in the present study on face processing (B). (A) The letter-string stimuli contained 19 categories. See Material and methods in Tarkiainen *et al.* (1999) for more details. (B) The face stimuli contained eight categories, which are demonstrated here by two examples from each category. The numbers incorporated in the names of the FACES categories refer to the noise level and the abbreviation PH denotes photographed images. The names shown above the images are used in the text to refer to the different stimulus types.

Puce *et al.*, 1996) suggests that these stimulus-specific signals may represent a more general phase of object-level analysis (Malach *et al.*, 1995).

Our aim in the present study was to identify the locus in time and space at which processing streams start to differ for faces and letter-strings. An appropriate paradigm was provided by our previous work on the visual processing of letter-strings (Tarkiainen *et al.*, 1999). Using letter- and symbol-strings of different length and variable levels of masking visual noise, we had succeeded in separating the processes of low-level visual analysis at 100 ms, which apparently reacts only to the visual complexity of images, and stimulus-specific processing at 150 ms after stimulus onset. In the present study, the same masking algorithm was employed to investigate processing of simple, drawn cartoon-like faces in 10 individuals who had also participated in the previous study on letter-strings. The results of the present study on faces were compared with our earlier findings on letter-string processing (Tarkiainen *et al.*, 1999) to reveal similarities and differences in brain activations in early face and letter-string processing.

Two hypotheses were tested in this study. (i) If the early response at 100 ms after stimulus reflects stimulus-non-specific visual feature analysis, it should be essentially identical for faces and letter-strings masked with noise. (ii) If

stimulus-specific object-level processing takes place in the inferior occipitotemporal cortex, we would expect dissociation into face- and letter-string-specific processing routes by the subsequent response at ~150 ms. Testing these hypotheses should allow us to establish the spatiotemporal limits of visual feature versus object-level analysis.

Material and methods

Subjects

Ten subjects took part in the present study. They had all participated in our earlier study of letter-string reading (Tarkiainen *et al.*, 1999). They were healthy, right-handed, Finnish adults (four females, six males, aged 23–44 years, mean age 31 years) and they gave their informed consent to participation in this study. The subjects were all university students or graduates, and their visual acuity was normal or corrected to normal.

Stimuli

The letter-string stimuli in Tarkiainen *et al.* (1999) contained 19 stimulus categories (Fig. 1A): letter-strings with three lengths (single letters, legitimate two-letter Finnish syllables, legitimate four-letter Finnish words), each with four levels of

masking noise (levels 0, 8, 16, 24; see below); symbol-strings of the same length (always with noise level 0, i.e. without noise) and empty noise patches (four levels, no letters or symbols).

The different noise levels used in the masking varied the visibility of the letter-strings. Noise was added to the originally noiseless images by changing the grey level of each pixel randomly. The amount of change was picked from a Gaussian distribution with zero mean and a standard deviation corresponding to the level of noise. If the new grey level was not within the possible range (0–63 from black to white), the procedure was repeated for that pixel. The addition of the Gaussian noise increased the local luminance contrast in the images, thus giving them a more complex appearance.

The face-object stimuli contained eight image categories (Fig. 1B). We constructed the stimuli with the intention of making measurement as similar as possible to that used in the letter-string study. The first stimulus category (FACES_0) was a collection of simple, cartoon-like, drawn faces with nine different expressions. The task during measurement was to identify the expressions and name them if asked, which forced the subject to analyse the images in a manner similar to reading the letter-strings, as described by Tarkiainen *et al.* (1999).

The second (FACES_8), third (FACES_16) and fourth (FACES_24) image categories consisted of the same drawn face images as those in the first category but the drawings were masked by an increasing level (levels 8, 16 and 24, respectively) of random Gaussian noise, making the recognition of the expressions harder. The grey levels of the drawn images and the noise levels masking the faces matched the values used in the letter-string study.

In the measurement of letter-strings, strings of geometrical symbols had served as control stimuli for letter-strings. We now included two control categories: the fifth stimulus category (OBJECTS) consisted of simple drawn images of common household objects (20 objects, e.g. a chair, a book, a hat, a mug) and the sixth category (MIXED) comprised eight images in which parts of the drawn faces were mixed (shifted and rotated) in a random manner and placed within geometrical objects to yield images that were as complex but not as readily recognized as those in FACES_0. Contrasting the FACES_0 and MIXED and OBJECTS conditions should provide information about the image attributes required by the face-specific neurones.

An additional comparison was enabled by the use of photographs. The seventh category (PH_FACES) consisted of black-and-white photographs of faces with expressions similar to those in the first category. The same male volunteer was viewed from the front in all face photographs; he was unknown to all our subjects. The last (eighth) category (PH_OBJECTS) was a collection of black-and-white photographs of objects similar to those in the fifth category.

Magnetoencephalography (MEG)

Brain activity was recorded with a 306-channel Vectorview magnetometer (Neuromag, Helsinki, Finland), which detects the weak magnetic fields created by synchronous activation of thousands of neurones. From the magnetic field patterns, we estimated the underlying cortical activations and their time courses of activation with equivalent current dipoles, as described by Tarkiainen *et al.* (1999). For a comprehensive review of MEG, see Hämäläinen *et al.* (1993).

Procedure

To reach an acceptable signal-to-noise ratio, the subject's brain responses must be averaged over several presentations of images belonging to each stimulus category. In the present study, we prepared one stimulus sequence in which all the different stimulus types (eight categories) appeared in pseudorandomized order but with equal probability. The only restriction was that the same image was not allowed to appear twice in a row. This sequence was presented to the subject in four shorter parts, each lasting ~8 min. Between different parts of the sequence, there were short breaks of 1–2 min to allow the subject to rest.

The subject's task during the MEG measurement was to pay attention to the stimuli and, when prompted by the appearance of a question mark, to read out loud the letter-string [this was done only in the previous study (Tarkiainen *et al.*, 1999)] or to say the name of the facial expression (present study) that was shown immediately before the question mark. No correct or desired names for different expressions were given, and it was emphasized that the subject should say the name that first comes to his or her mind. The purpose of showing the question mark was to ensure that the subject stayed alert. The question mark appeared with the probability of 1.5% in both the letter-string and the face study.

The MEG measurements took place in a magnetically shielded room (Euroshield Oy, Eura, Finland), where the subject sat with his or her head resting against the measurement helmet. The room was dimly lit and the images were presented on a rear projection screen with a data projector (Vista Pro; Electrohome, Kitchener, Ontario, Canada) controlled by a Macintosh Quadra 800 computer. The projection screen was placed in front of the subject at a distance of ~90 cm. All the images were presented at the same location on the screen at a comfortable central viewing position. The letter-string stimuli used by Tarkiainen *et al.* (1999) occupied a visual angle of $\sim 5^\circ \times 2^\circ$ and the face images in the present study a visual angle of $\sim 5^\circ \times 5^\circ$. Subjects were asked to fixate the central part of the screen, where the images appeared. The images were shown on a grey background. The grey level matched the mean grey level of the stimulus images and was used to keep the luminance relatively constant and to reduce the stress to the eyes caused by viewing the stimuli. With our visual stimulus presentation hardware, the actual stimulus

image appeared 33 ms later than the computer-generated trigger that marked the onset of the stimulus. This delay was taken into account in the results, and the latencies refer to the appearance of the stimuli on the screen.

The anatomical information given by the subjects' MRIs was aligned with the coordinate system of the MEG measurement by defining a head coordinate system with three anatomical landmarks (the nasion and points immediately anterior to the ear canals). Four small coils attached to the subject's head allowed the measurement of the position of the head with respect to the MEG helmet. Active brain areas were localized by means of the head coordinate system, in which the x -axis runs from left to right through points anterior to the ear canals, the y -axis towards the nasion and the z -axis towards the top of the head.

During a recording session in the present study, the stimulus images appeared for 100 ms with a 2-s interstimulus interval. The only exception was the question mark, which was shown for 2 s to allow the subject to name the facial expression. The 60-ms stimulus presentation time that was used in the letter-string study (Tarkiainen *et al.*, 1999) was also tested, but we found that it was not always long enough to allow the recognition of expressions, especially from the photographs.

MEG signals were pass-band filtered at 0.1–200 Hz and sampled at 600 Hz. Signals were averaged over an interval starting 0.2 s before and ending 0.8 s after the onset of the image. The vertical and horizontal electro-oculograms were monitored continuously and epochs contaminated by eye movements and blinks were excluded from the averages. The smallest number of averages collected in one category was 87. On average, 102 epochs were averaged for each stimulus category. After the MEG sessions, to check for intersubject consistency in naming the expressions, the subjects were asked to name the different expressions in writing.

Data analysis

Averaged MEG responses were low-pass filtered digitally at 40 Hz. The baseline for the signal value was calculated from a time interval of 200 ms before image onset. Vectorview employs 204 planar gradiometers and 102 magnetometers, with two orthogonally oriented gradiometers and one magnetometer at each of the 102 measurement locations. Magnetometers may detect activity from deeper brain structures than gradiometers, but they are also more sensitive to external noise sources. Since we were interested in cortical activations, we based our analysis on the gradiometer data only.

Activated brain areas and their time courses were determined using equivalent current dipoles. For this purpose, each subject's brain was modelled as a sphere matching the local curvature of the occipital and temporal regions. In individual subjects, the MEG signals obtained within 700 ms after image onset were analysed in all stimulus conditions, and the separately determined current dipoles were combined into a

single multidipole model of nine to 14 dipoles (mean number 11), which accounted for the activation patterns in all stimulus conditions. By applying the individual multidipole models to the different stimulus conditions, we obtained the amplitude waveforms of different source areas. On the basis of these amplitude waveforms, we identified the sources that showed systematic stimulus-dependent behaviour. This procedure is explained in detail in the Results section.

Statistical tests on the amplitude and latency behaviour of selected source groups were carried out by ANOVA (analysis of variance) and the t -test. To avoid any bias that might result from sources that were not statistically independent, only one source per subject was accepted for these tests. If a subject had multiple sources belonging to the same source group, the mean value was used for the amplitude tests. The latency tests were performed with the source showing the shortest latency, but only if it had a clear activation peak in all stimulus conditions.

For visualization purposes, all the individual source coordinates were transformed into standard brain coordinates. This alignment was based on a 12-parameter affine transformation (Woods *et al.*, 1998) followed by a refinement with a non-linear elastic transformation (Schormann *et al.*, 1996), in which each individual brain was matched to the standard brain by comparing the greyscale values of the MRIs. All the source location parameters reported in the Results section were calculated from the transformed coordinates.

Results

Summary of activation patterns observed in the letter-string study

The main findings of our earlier study of letter-string processing (Tarkiainen *et al.*, 1999) are illustrated in Fig. 2, which shows the responses to one-item (single letters or symbols) and four-item (four-letter words or symbols) strings for the 10 individuals who also participated in the present study of face processing (data on empty noise patches and two-item strings are not shown). Two distinct patterns of activation were identified within 200 ms after stimulus onset. The first pattern, named Type I, took place ~100 ms after image onset. These responses originated in the occipital cortex close to the midline (15 sources from eight subjects). They were not specific to the type of string, as letter-strings and symbol-strings evoked equally strong responses. However, Type I activation reacted strongly to the visual complexity of the images, and the strongest responses measured were the responses to images with the highest amount of visual features (noise, number of items).

The second pattern of activation, named Type II, was seen ~150 ms after image onset. These responses originated in the inferior occipitotemporal region with LH dominance (10 LH and three RH sources from 10 subjects; LH sources were found in nine subjects and RH sources in three subjects). They were letter-string-specific in the sense that responses

were stronger for letter-strings than symbol-strings and they collapsed for the highest level of noise-masking (contrary to Type I activation).

Type I activation for faces

In our study of letter-strings, we classified all early (<130 ms) sources showing a systematic increase in amplitude with the level of noise as the noise-sensitive Type I activation group. In the present study, the stimuli did not include the empty noise patches that were used in the selection procedure in the letter-string study (Tarkiainen *et al.*, 1999), but the same noise levels were used to mask drawn faces (categories 1–4, representing noise levels 0, 8, 16 and 24, respectively). Thus, the selection was now based on the comparison of noiseless drawn faces (FACES_0) versus faces masked with the highest

level of noise (FACES_24). As only one comparison was used, we set a strict requirement for a significant difference. Only those sources were selected for which the peak activation in response to heavy noise (FACES_24) was stronger than to noiseless drawn faces (FACES_0) by at least 3.29 times the baseline standard deviation (corresponding to $P < 0.001$). Exactly as in the letter-string study, the upper time limit for activation peaks was set to 130 ms (in the FACES_24 condition). According to these criteria, we accepted 24 sources from nine subjects. The only subject who did not show any Type I sources did not show them in the earlier study (Tarkiainen *et al.*, 1999) either. One subject who had very strong and widespread occipital activity had as many as seven Type I sources. The other subjects had on average two Type I sources.

The Type I source locations collected from all subjects are shown in Fig. 3A. Sources were located bilaterally in the occipital region with a distance (mean \pm SEM) of 21 ± 3 mm from the occipital midline. The behaviour of all Type I source areas in different stimulus conditions is summarized in Fig. 3B and C. Figure 3B shows the Type I peak amplitudes, which were first normalized with respect to the FACES_24 condition [equal to 1; strength (mean \pm SEM) 33 ± 5 nAm] and then averaged over all nine subjects (24 sources in total). If no clear peak was found for a condition (usually for some of the noiseless conditions), the baseline standard deviation was used as the amplitude and no peak latency was obtained for that situation.

Similar to the findings in Tarkiainen *et al.* (1999), the activation strengths of these sources increased systematically with the level of noise. The effect of noise on Type I amplitudes was significant for drawn faces [repeated measures

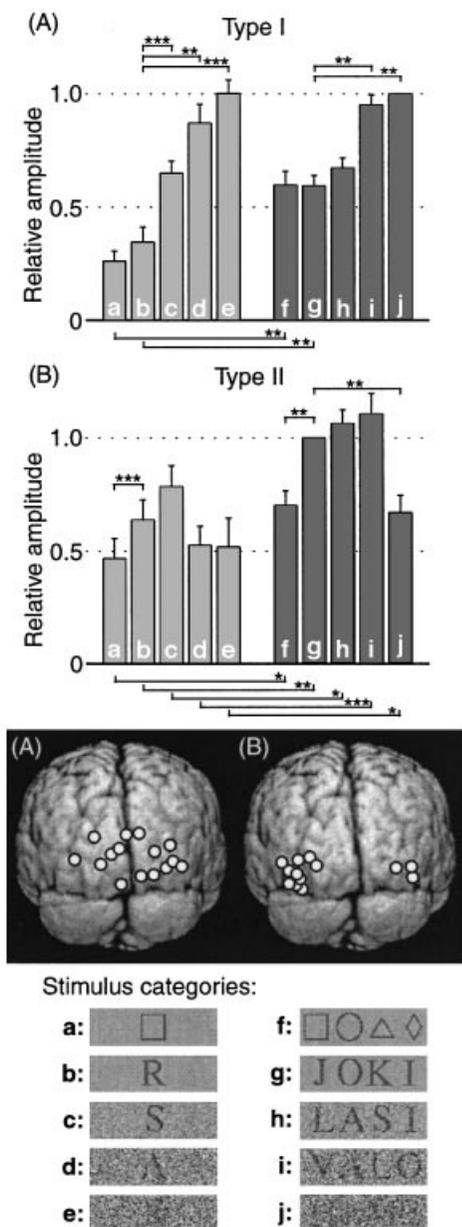


Fig. 2 The main results of the earlier letter-string study (for a full account, see Tarkiainen *et al.*, 1999) were recalculated for the 10 subjects who participated in the present study. Mean (+ standard error of the mean) amplitude behaviour is shown for the single letter/symbol (a to e) and four-letters/symbols (f to j) conditions. (A) Occipital Type I responses reached their maximum ~100 ms after image onset and increased with the level of noise and string length. Individual source amplitudes were scaled with respect to the noisiest word condition (j) and were averaged across all sources (15 sources from eight subjects). (B) Occipitotemporal Type II responses showed specificity for letter-strings ~150 ms after image onset. Individual source amplitudes were scaled with respect to the visible words condition (g) and averaged across all sources (13 sources from 10 subjects). Differences in the activation strengths in (A) and (B) were tested with paired *t*-tests (calculated from absolute amplitudes) for the following pairs: (i) noiseless letter- versus symbol-strings (a versus b and f versus g); (ii) noiseless versus noisy letter-strings (b versus c, d, e, and g versus h, i, j); and (iii) all corresponding one-item versus four-item strings (a versus f, b versus g, c versus h, d versus i, and e versus j). Significant differences are marked in the figures: * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. The centre points of the source areas are indicated as dots on an MRI surface rendering of the standard brain geometry. The brain is viewed from the back and the sources are projected on the brain surface for easy visualization.

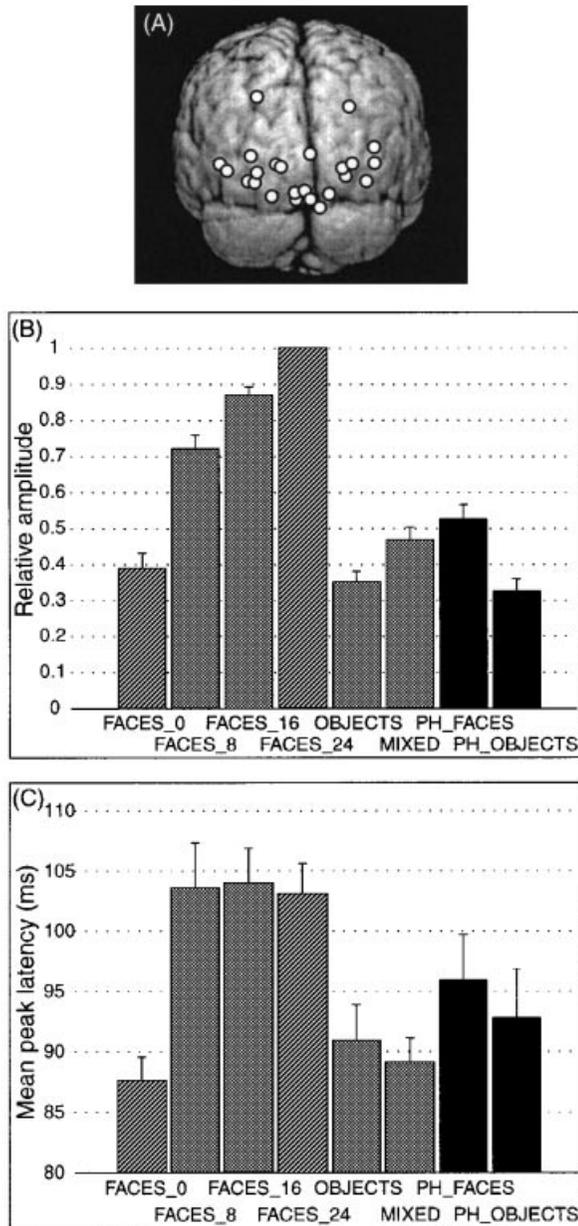


Fig. 3 (A) Locations of all Type I_f sources (Type I sources found in the present study) on an MRI surface rendering of the standard brain geometry. The brain is viewed from the back. (B) Mean (+ standard error of the mean) Type I_f source amplitudes are shown relative to the FACES₂₄ condition (set equal to 1). (C) Mean (+ standard error of the mean) Type I_f source peak latencies. The values are calculated across all Type I_f sources (24 sources from nine subjects).

ANOVA with image type (FACES_{0/8/16/24}) as a within-subjects factor, $F(3,24) = 15.0$, $P = 0.00001$, calculated from absolute amplitudes]. Type I activation strengths also differed among the noiseless image categories [repeated measures ANOVA with image type (FACES₀, PH_FACES, OBJECTS, PH_OBJECTS, MIXED) as a within-subjects factor, $F(4,32) = 6.5$, $P < 0.001$]. Pairwise comparisons showed that the amplitude difference between photographed

faces (PH_FACES) and objects (PH_OBJECTS) was significant ($P < 0.01$, paired two-tailed t -test) but the difference between FACES₀ and PH_FACES and that between FACES₀ and OBJECTS were not significant. The effect of noise on Type I amplitudes was very clear as even the difference between noise level 8 (FACES₈) and all noiseless image types was significant ($P < 0.05$ for all paired two-tailed t -tests).

In the peak latencies (Fig. 3C), only one phenomenon was evident. The latencies were shorter for noiseless conditions than for noisy images [repeated measures ANOVA with image type, all categories, as a within-subjects factor, $F(7,42) = 3.5$, $P < 0.01$; only the earliest source in FACES₂₄ condition was included]. However, the onset latencies showed no difference between FACES₀ and FACES₂₄ (paired two-tailed t -test). The apparent difference in peak latencies therefore arises from the fact that the responses to noiseless stimuli were smaller in amplitude and thus reached the maximum earlier than responses to noisy stimuli. The peak latency of all Type I sources for FACES₂₄ (mean \pm SEM) was 103 ± 3 ms.

Comparison of Type I activity between word and face processing

The behaviour of Type I sources in the present study on faces (henceforth referred to as Type I_f sources) was very similar to that found in our earlier study of letter-strings (Tarkiainen *et al.*, 1999) (henceforth referred to as Type I_s sources).

The peak latencies measured for faces were identical to those measured in the letter-string reading task, namely 105 ± 4 ms for the four-letter words masked with the highest level of noise and 103 ± 3 ms for the drawn faces masked with the highest level of noise (FACES₂₄). The onset latencies were also indistinguishable (69 ± 3 and 64 ± 3 ms for the words with noise level 24 and FACES₂₄, respectively). However, the amplitudes differed clearly. The mean peak amplitude for FACES₂₄ was 33 ± 5 nAm, whereas for four-letter words with noise level 24 the mean value was only 14 ± 3 nAm.

The source locations were, on average, similar between the two studies. The sources were located mainly in the visual cortices surrounding the V1 cortex and distributed along the ventral visual stream.

When only the earliest Type I source was selected for each subject, the locations could be compared at the individual level. The earliest Type I_f and Type I_s sources were typically not located at the same exact position in the same subject but were separated on average by 24 ± 6 mm (averaged over all subjects showing Type I behaviour). However, paired t -tests showed no systematic differences between Type I_f and Type I_s source locations along any coordinate axis. The total number of Type I sources was higher in the present study (24 sources) than in our earlier study (15 sources for these 10 subjects).

The differences in the number and strength of Type I sources between letter-string and face processing are prob-

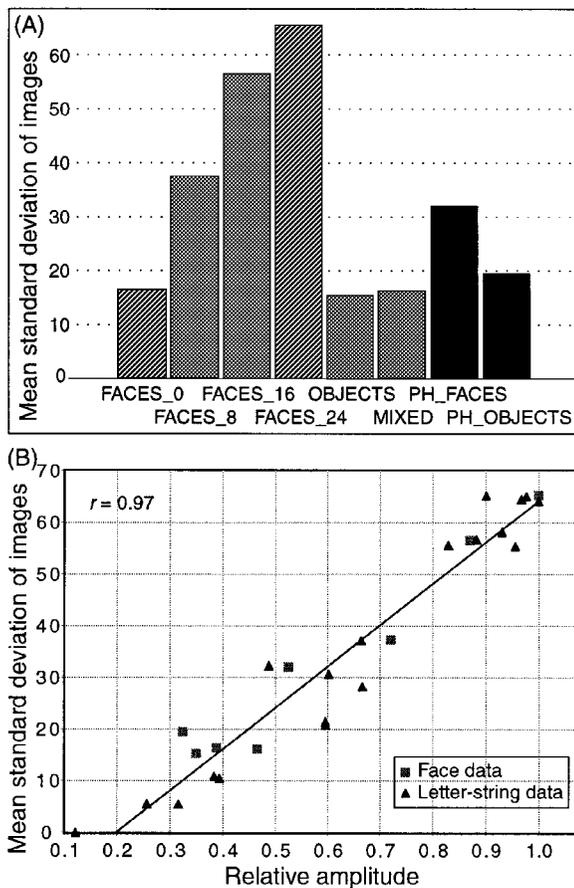


Fig. 4 (A) Mean standard deviation (for details, see Results) of the greyscale values (0–255) of each stimulus category. Note the similarity to Fig. 3B. (B) Correlation between the mean standard deviation of image categories and the Type I mean relative amplitudes is calculated for all the 19 stimulus categories of the letter-string study (triangles) and for the eight stimulus categories of the present study (squares).

ably caused mainly by differences in stimulus size and in the visual presentation hardware that affected the luminance of images. In addition, the calibration of the Vectorview MEG system used in the present study is different from that of the Neuromag-122 system used in the earlier, letter-string experiment. The measurement array of the Vectorview system also covers better the lower occipital areas, which may have enabled us to detect source areas not as readily accessible with the Neuromag-122 device. The slightly modified selection criteria for Type I activation may also have generated small differences in the results. Therefore, we do not consider the differences in source number or strength as important and they will not be discussed further.

Relationship of Type I activity to visual complexity of images

Type I_s activity increased with the level of noise as well as with the length of the string. A clear increase with the amount

of masking noise was also evident in Type I_f activity, whereas only small differences were seen between all the noiseless image types. The most important factor affecting Type I activation may thus be the visual complexity of the image. To test this hypothesis, we defined the complexity of our stimulus images in the following way. Each image was represented by an $m \times n$ matrix, where m is the height of the image and n is the width of the image in pixels and each matrix element gives the greyscale value of the corresponding pixel. For all image matrices belonging to the same stimulus category, we calculated the column-wise (could equally have been row-wise) standard deviations of greyscale values and used the mean value to represent that stimulus category. The mean standard deviations calculated in this way for the face and object stimuli are shown in Fig. 4A. This result resembled closely the source strength behaviour seen in Fig. 3B. As illustrated in Fig. 4B, we obtained a strong correlation ($r = 0.97$, $P < 0.00001$) between the mean peak amplitudes of Type I sources (averaged over all subjects) and the mean standard deviations of the corresponding stimulus images (averaged over all images belonging to the same category) when the results from both the present study (eight stimulus categories) and the letter-string study (19 stimulus categories; Tarkiainen *et al.*, 1999) were combined.

Type II activation for faces

In our study of reading, the letter-string-specific Type II sources were selected by comparing the activations between four-item letter- and symbol-strings (Tarkiainen *et al.*, 1999). In the present study, we made a similar comparison between face and object categories. Sources were included in the face-specific Type II category when amplitude waveforms peaked after the Type I activation of the same subject but before 200 ms (in FACES_0), and when peak amplitudes for faces exceeded those for objects (both in drawn and photographed form) by at least 1.96 times the baseline standard deviation (corresponding to $P < 0.05$). Thus, our definition of ‘face-specificity’ does not mean activation only for face images but activation that is clearly stronger for faces than for objects.

Sources that fulfilled these criteria were found in all 10 of our subjects, and a total of 19 sources (one to three per subject) were classified as Type II. Eleven of these sources were located in the right inferior occipitotemporal area, one in the RH but close to the occipital midline, and seven in the left inferior occipitotemporal cortex (Fig. 5A). All 10 subjects had at least one RH Type II source and seven subjects had also one additional LH source.

Figure 5B shows the average amplitude behaviour of all Type II source areas scaled with respect to FACES_0 (mean \pm SEM amplitude 34 ± 4 nAm). As expected, the differences between faces and other noiseless image types were clear [repeated measures ANOVA with image type (FACES_0, PH_FACES, OBJECTS, PH_OBJECTS, MIXED) as a within-subjects factor, $F(4,36) = 27.4$, $P < 0.00001$, calculated from absolute amplitudes]. The effect of noise was also

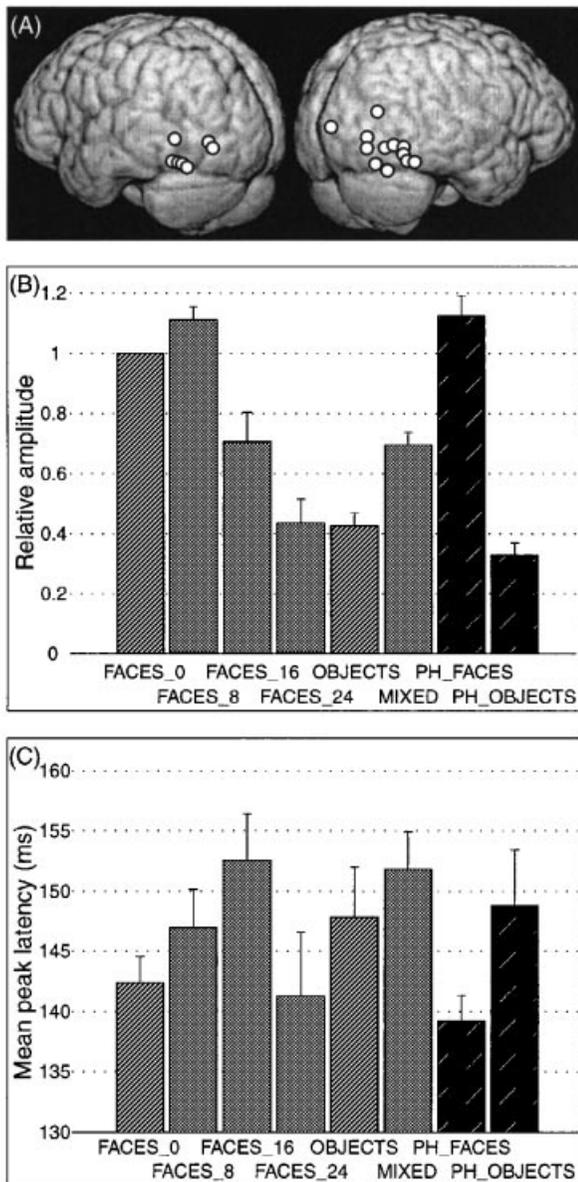


Fig. 5 (A) Locations of all Type II_f sources presented on an MRI surface rendering of the standard brain geometry. The brain is viewed from the back but rotated slightly to the left and right. (B) Mean (+ SEM) Type II_f source amplitudes are shown relative to the FACES₀ condition (set equal to 1). (C) Mean (+ SEM) Type II_f source peak latencies. The values are calculated across all Type II_f sources (19 sources from 10 subjects).

significant [repeated measures ANOVA with image type (FACES_{0/8/16/24}) as a within-subjects factor, $F(3,27) = 19.0$, $P < 0.00001$]. The activation strength of Type II sources increased slightly with low noise (level 8), but collapsed for the highest noise level.

Interestingly, MIXED faces evoked stronger activation than OBJECTS ($P < 0.01$, paired two-tailed t -test) but weaker activation than FACES₀ ($P < 0.01$). Activation strengths of FACES₀ and PH_FACES did not differ.

The mean peak latencies (Fig. 5C) were not significantly different among the conditions [repeated measures ANOVA with stimulus type, all categories, as a within-subjects factor, $F(7,35) = 2.0$, $P = 0.08$; only the earliest source in FACES₀ condition was included], but pairwise comparisons revealed some differences. The apparently longer latencies for objects than for faces reached significance only for the photographic images (PH_FACES versus PH_OBJECTS, $P < 0.05$, paired two-tailed t -test). Responses to MIXED faces were also significantly delayed with respect to those to FACES₀ ($P = 0.001$). The peak latency of all Type II sources for FACES₀ (mean \pm SEM) was 142 ± 3 ms.

In Fig. 5, the results are pooled over the LH and RH Type II sources. Hemispheric comparison was only possible for those seven subjects who had a Type II source in both the left and the right occipitotemporal cortex. The main effect of hemisphere was not significant [repeated measures ANOVA with hemisphere (RH, LH) and image type, all categories, as within-subjects factors, $F(1,6) = 1.1$, $P = 0.3$] but the two-way interaction of hemisphere \times image type reached significance [$F(7,42) = 2.6$, $P < 0.05$]. Pairwise comparisons revealed a significant difference between LH and RH activation only in the FACES₂₄ condition ($P < 0.01$, paired two-tailed t -test), in which LH activation (16 ± 3 nAm) was stronger than RH (8 ± 3 nAm) activation. The mean onset and peak latencies in the FACES₀ condition were 110 ± 4 and 144 ± 3 ms, respectively, in the right occipitotemporal cortex and 101 ± 5 and 142 ± 4 ms in the left occipitotemporal cortex. The mean distance of all Type II sources from the midline was 33 ± 3 mm. The LH and RH sources were located symmetrically with respect to the head midline.

Comparison of Type II activity between word and face processing

The hemispheric distribution of face-specific (12 RH and seven LH) and letter-string-specific (three RH and 10 LH) sources was different ($P < 0.05$, Fisher's exact probability test). The letter-string-specific Type II_{ls} sources expressed LH dominance ($P < 0.05$, binomial test), whereas the face-specific Type II_f sources were rather found bilaterally and their number suggested only slight preference for the RH.

When the hemispheric distribution was ignored, the locations of source areas in the inferior occipitotemporal cortex were very similar. The mean distance from the midline was 34 ± 2 mm for both Type II_{ls} and Type II_f sources (the one midline RH Type II_f source was excluded). In the inferior–superior (z -coordinate) direction, the mean value for Type II_{ls} and Type II_f sources was 37 ± 2 and 37 ± 3 mm, respectively. The only small difference was seen in the anterior–posterior (y -coordinate) direction, where the mean values for Type II_{ls} and Type II_f sources were -47 ± 2 and -41 ± 2 mm, respectively, indicating that the centre of activation of Type II_f sources was located on average 6 mm anterior to the centre of activation of Type II_{ls} sources. This

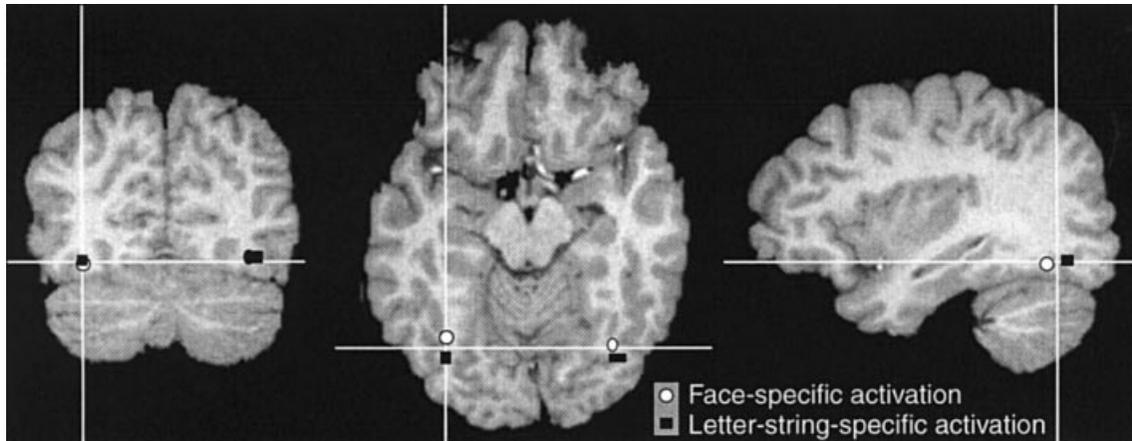


Fig. 6 Mean locations of the left- and right-hemisphere Type II sources presented on (from left to right) coronal, horizontal and sagittal (left hemisphere only) slices of the standard brain MRIs. White lines denote the relative locations of the slices. The mean locations of the face-specific Type II_f sources are marked with white ellipses and the mean locations of the letter-string-specific Type II_{ls} sources with black rectangles. The dimensions (axes) of ellipses and rectangles are equal to twice the standard error of the mean. The only clearly deviant (close to midline) Type II_f source was excluded from the calculation of the right-hemisphere Type II_f mean location. In coronal and horizontal slices, left is on the left.

difference reached significance in a two-tailed *t*-test when all the Type II_{ls} and Type II_f source locations were considered ($P < 0.05$), but not when only LH source locations were considered (in five out of six individuals the Type II_f source was anterior to the Type II_{ls} source). The mean locations of Type II sources in Talairach coordinates (Talairach and Tournoux, 1988) were (x, y, z) $-37, -70, -12$ and $35, -73, -10$ mm for LH and RH Type II_{ls} sources, respectively, and $-37, -61, -13$ and $33, -68, -10$ mm for LH and RH Type II_f sources. Fig. 6 shows the mean locations of Type II_{ls} and Type II_f sources in a standard brain.

The timing of Type II activity in face and letter-string processing was identical. The mean peak latency of all Type II_{ls} sources for four-letter words without noise was 141 ± 4 ms and that for Type II_f sources in the FACES_0 condition was 142 ± 3 ms. No difference was seen in the onset latencies (108 ± 5 ms and 106 ± 3 ms, respectively) either. The activation was again stronger in the present data set. The mean peak amplitude for Type II_{ls} sources for four-letter words without noise was 17 ± 2 nAm versus 34 ± 4 nAm for Type II_f sources in the FACES_0 condition.

Discussion

Our results show that the early visual processing of faces and letter-strings (summarized in Fig. 7) consists of at least two distinguishable processes taking place in the occipital and occipitotemporal cortices within 200 ms after stimulus presentation. The first process, which we have named Type I, took place ~ 100 ms after image onset in areas surrounding V1 cortex. This activity was not sensitive to the specific content of the stimulus and was common to the processing of both letter-strings and faces. Type I sources showed a

monotonic increase in signal strength as a function of the visual complexity of the images. Some 30–50 ms after Type I activation and ~ 150 ms after stimulus onset, stimulus-specific activation (Type II) emerged in the inferior occipitotemporal cortices. Although both letter-strings and faces activated largely overlapping areas in the inferior occipitotemporal cortex, the hemispheric distribution of these areas was different. Letter-string processing was concentrated in the LH, whereas face processing occurred more bilaterally, apparently with slight RH dominance.

Visual feature analysis

The combined letter-string and face data revealed that Type I activation strength correlates well with the simple image parameter of the mean standard deviation of greyscale values, providing strong support for the interpretation that Type I activation is related to low-level visual feature analysis, such as the extraction of oriented contrast borders. Spatially and temporally similar increased activation for scrambled images has also been reported, e.g. by Allison *et al.* (1994, 1999), Bentin *et al.* (1996) and Halgren *et al.* (2000).

Both for faces and letter-strings, Type I sources were located in the occipital cortex close to the midline. At the individual level, some differences were seen in the Type I source locations. This, however, was not surprising considering the differences in the stimulus presentation and measurement hardware between the two studies. The mean location parameters matched well between face and letter-string measurements. The timing of Type I activation was identical in both measurements. Type I activation elicited by faces is thus essentially similar to the activation evoked by letter-strings.

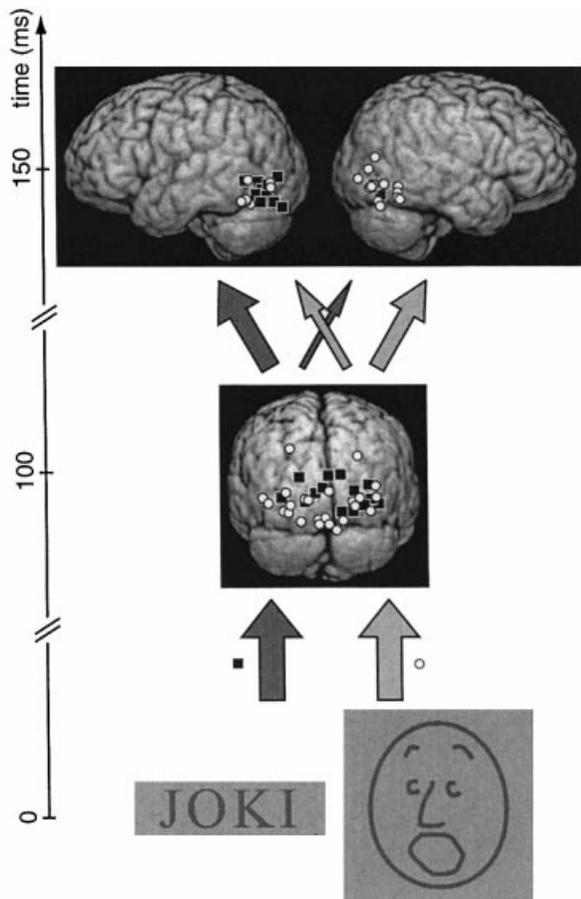


Fig. 7 Early (<200 ms) processing of face and letter-string information consists of at least two distinct stages. The middle part of the figure illustrates the first stage (Type I), which takes place in the occipital cortex ~100 ms after image onset. This activation does not differ between face and letter-string processing. The locations of Type I sources evoked by letter-strings are marked with black squares and the locations of Type I sources activated by faces with white circles. The second activation pattern ~150 ms after image onset is specific to the stimulus-type (top part of the figure), with strong lateralization to the left hemisphere for letter-strings (black squares) and slight right-hemisphere preponderance for faces (white circles). See Discussion for details. Sources are gathered from all 10 subjects and their locations are shown on MRI surface renderings of the standard brain geometry.

Object-level processing

Face-specific Type II sources were located in the inferior occipitotemporal cortices bilaterally, even though slightly more sources were located in the RH. These results are in good accordance with the results of Sergent *et al.* (1992), Haxby *et al.* (1994), Kanwisher *et al.* (1997), Gorno-Tempini *et al.* (1998), Wojciulik *et al.* (1998), Allison *et al.* (1999) and Halgren *et al.* (2000). If the hemispheric distribution is ignored, the locations of both face- and letter-string-specific Type II sources were very similar. The mean location parameters differed only in the anterior–posterior direction as the centres of face-specific source areas were on average

6 mm more anterior than centres of letter-string-specific source areas. This difference is small and, taking into account the changes in the stimulus and measurement hardware, perhaps unimportant. However, we cannot exclude the possibility that differences in Type II source locations reflect the spatial separation of functionally distinct neural systems in the occipitotemporal cortex. Interestingly, on the basis of intracranial event-related potential (ERP) studies, Puce *et al.* (1999) reported that face-specific sites were typically anterior to letter-specific sites in the occipitotemporal cortex.

More than anything else, we want to stress the extreme similarity of letter-string- and face-specific occipitotemporal activations. Even though the hemispheric balance was different, the activated areas within the inferior occipitotemporal cortex were very close to each other. Also, the timing of these two activations was practically identical. This is noteworthy, since the visual properties of faces are quite different from those of letter-strings and, importantly, since reading and face processing also differ on an evolutionary scale. Still, it seems that both skills use highly similar cortical systems within the visual domain. We conjecture that the inferior occipitotemporal activations at ~150 ms after stimulus onset represent a more general object-level processing stage that takes place after the common low-level analysis of visual features and acts as a gateway to higher processing areas. In our studies, this activation was seen with two different classes of visual stimuli united by their importance to the modern-day human, namely faces and letter-strings. As the ability to recognize letter-strings accurately and quickly develops through practice, it is possible that similar abilities, if needed, can also be developed and their signature detected at the cortical level for other classes of objects. This is demonstrated by Allison *et al.* (1994), who identified responses specific to Arabic numbers in areas close to those responding specifically to faces and letter-strings, and by Gauthier *et al.* (1999, 2000), who showed that the face-specific fusiform area can be activated (in a manner similar to activation by faces) also by other classes of objects (a novel group called ‘greebles’, birds and cars) in individuals who have had a lot of practice in recognizing the objects in their field of expertise.

One striking feature in our results is the high level of occipitotemporal activation evoked by the very simple drawn images. Halgren *et al.* (2000) reported that schematic face sketches evoked ~30% less face-specific occipitotemporal activation than face photographs. We did not observe such a general difference (Fig. 5B), although our drawn images were simpler than those used by Halgren *et al.* (2000). A likely reason for this difference is the short stimulus presentation time we used, which may not always allow the full recognition and categorization of complex face photographs. The MIXED face images, which contained the same face components as drawn faces but in randomized positions and orientations and placed inside different geometrical shapes, evoked weaker and somewhat delayed activation when compared with normal face images. However, they still

activated the face-specific occipitotemporal areas more than fully meaningful images of familiar objects; this shows that even a few drawn lines that resemble parts of faces are enough to activate these areas and supports the notion of structural encoding of different face components (Eimer, 1998). All in all, our results once again demonstrate the amazing ability of the human brain to recognize faces that are very unnatural in their appearance.

One might be tempted to question the face-specificity of our Type II responses and explain it only as a consequence of our task, which directed the subject to pay more attention to the faces than to the other image categories. Wojciulik *et al.* (1998), using functional MRI, showed that the fusiform face activation can be modulated by voluntary attention. On the other hand, Puce *et al.* (1999) demonstrated that top-down influences (semantic priming and face-name learning and identification) did not affect the ventral face-specificity within 200 ms of image onset, but affected the responses measured from the same locations later in time. Even if attention plays a major role in these early object-specific responses, dissociation of the processing pathways obviously occurred ~150 ms after stimulus onset, with lateralization to the left inferior occipitotemporal cortex for attended letter-strings and a more bilateral response pattern to attended faces. Whether attention amplifies the lateralization remains to be answered by further studies.

In conclusion, the strong correlation of the noise-sensitive occipital activation at 100 ms with stimulus complexity, which was similar for letter-strings and faces, confirms the role of this response in low-level visual feature processing. The subsequent inferior occipitotemporal activation at 150 ms, similar in timing and location but different in hemispheric distribution for letter-strings and faces, apparently reflects the earliest stage of stimulus-specific object-level processing.

Acknowledgements

We wish to thank Mika Seppä for help in transforming the individual subject data to standard brain coordinates, Martin Tovee for volunteering to serve as our photographic model, Päivi Helenius for help in statistical analysis and comments on the manuscript and Minna Vihla for comments on the manuscript. This work was supported by the Academy of Finland (grant 32731), the Ministry of Education of Finland, the Human Frontier Science Program (grant RG82/1997-B), the Wellcome Trust and the European Union's Large-Scale Facility Neuro-BIRCH II at the Low Temperature Laboratory, Helsinki University of Technology.

References

Allison T, McCarthy G, Nobre A, Puce A, Belger A. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. [Review]. *Cereb Cortex* 1994; 4: 544–54.

Allison T, Puce A, Spencer DD, McCarthy G. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex* 1999; 9: 415–30.

Bentin S, Allison T, Puce A, Perez E, McCarthy G. Electrophysiological studies of face perception in humans. *J Cogn Neurosci* 1996; 8: 551–65.

Clark VP, Keil K, Maisog JM, Courtney S, Ungerleider LG, Haxby JV. Functional magnetic resonance imaging of human visual cortex during face matching: a comparison with positron emission tomography. *Neuroimage* 1996; 4: 1–15.

Damasio AR, Tranel D, Damasio H. Face agnosia and the neural substrates of memory. [Review]. *Annu Rev Neurosci* 1990; 13: 89–109.

De Renzi E, Perani D, Carlesimo GA, Silveri MC, Fazio F. Prosopagnosia can be associated with damage confined to the right hemisphere—an MRI and PET study and a review of the literature. [Review]. *Neuropsychologia* 1994; 32: 893–902.

Eimer M. Does the face-specific N170 component reflect the activity of a specialized eye processor? *Neuroreport* 1998; 9: 2945–8.

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci* 1999; 2: 568–73.

Gauthier I, Skudlarski P, Gore JC, Anderson AW. Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci* 2000; 3: 191–7.

Gorno-Tempini ML, Price CJ, Josephs O, Vandenberghe R, Cappa SF, Kapur N, et al. The neural systems sustaining face and proper-name processing. *Brain* 1998; 121: 2103–18.

Halgren E, Raji T, Marinkovic K, Jousmäki V, Hari R. Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb Cortex* 2000; 10: 69–81.

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. [Review]. *Rev Mod Phys* 1993; 65: 413–97.

Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J Neurosci* 1994; 14: 6336–53.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 1997; 17: 4302–11.

Kuriki S, Takeuchi F, Hirata Y. Neural processing of words in the human extrastriate visual cortex. *Brain Res Cogn Brain Res* 1998; 6: 193–203.

Lu ST, Hämäläinen MS, Hari R, Ilmoniemi RJ, Lounasmaa OV, Sams M, et al. Seeing faces activates three separate areas outside the occipital visual cortex in man. *Neuroscience* 1991; 43: 287–90.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, et al. Object-related activity revealed by functional magnetic

- resonance imaging in human occipital cortex. *Proc Natl Acad Sci USA* 1995; 92: 8135–9.
- McCarthy G, Puce A, Gore JC, Allison T. Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci* 1997; 9: 605–10.
- McCarthy G, Puce A, Belger A, Allison T. Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb Cortex* 1999; 9: 431–44.
- Nobre AC, Allison T, McCarthy G. Word recognition in the human inferior temporal lobe. *Nature* 1994; 372: 260–3.
- Puce A, Allison T, Asgari M, Gore JC, McCarthy G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci* 1996; 16: 5205–15.
- Puce A, Allison T, McCarthy G. Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb Cortex* 1999; 9: 445–58.
- Salmelin R, Service E, Kiesilä P, Uutela K, Salonen O. Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Ann Neurol* 1996; 40: 157–62.
- Sams M, Hietanen JK, Hari R, Ilmoniemi RJ, Lounasmaa OV. Face-specific responses from the human inferior occipito-temporal cortex. *Neuroscience* 1997; 77: 49–55.
- Schormann T, Henn S, Zilles K. A new approach to fast elastic alignment with applications to human brains. In: Höhne KH, Kikinis R, editors. *Visualization in biomedical computing. Lecture notes in computer science*, Vol. 1131. Berlin: Springer; 1996. p. 337–42.
- Sergent J, Ohta S, MacDonald B. Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain* 1992; 115: 15–36.
- Swithey SJ, Bailey AJ, Bräutigam S, Josephs OE, Jousmäki V, Tesche CD. Neural processing of human faces: a magnetoencephalographic study. *Exp Brain Res* 1998; 118: 501–10.
- Talairach J, Tournoux P. *Co-planar stereotaxic atlas of the human brain*. Stuttgart: Thieme Medical; 1988.
- Tarkiainen A, Helenius P, Hansen PC, Cornelissen PL, Salmelin R. Dynamics of letter string perception in the human occipitotemporal cortex. *Brain* 1999; 122: 2119–32.
- Wojciulik E, Kanwisher N, Driver J. Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *J Neurophysiol* 1998; 79: 1574–8.
- Woods RP, Grafton ST, Watson JD, Sicotte NL, Mazziotta JC. Automated image registration: II. Intersubject validation of linear and nonlinear models. *J Comput Assist Tomogr* 1998; 22: 153–65.

Received July 9, 2001. Revised December 3, 2001.

Accepted January 3, 2002