A principled method for determining the functionality of brain responses

Philippe G. Schyns,[CA] Ines Jentzsch, Mark Johnson,[1] Stefan R. Schweinberger and Frédéric Gosselin[2]

Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB; [1]School of Psychology, Birkbeck College, Malet Street, London WC1E 7HX, UK; [2]Départment de Psychologie, Université de Montréal, CP 6128, succ. Centre-ville, Montréal H3C 3J7, Canada

[CA]Corresponding Author: philippe@psy.gla.ac.uk

A challenging issue in relating brain function to perception and cognition concerns the functional interpretation of brain responses. For example, while there is agreement that the N170 component of event-related potentials is sensitive to face processing, there is considerable debate about whether its response reflects a structural encoder for faces, a feature (e.g. eye) detector, or something else. We introduce a principled approach to determine the stimulus features driving brain responses. Our analyses on two observers resolving different face categorization tasks (gender and expressive or not) reveal that the N170 responds to the eyes within a face irrespective of task demands. This suggests a new methodology to attribute function to different components of the neural system for perceiving complex stimuli. *NeuroReport* 14:1665–1669 © 2003 Lippincott Williams & Wilkins.

Key words: Attention; Brain potentials; Face categorization; Mechanisms; Perception; Recognition; Selective attention

INTRODUCTION

Face recognition has long been known to be of tremendous importance for the normal social functioning of humans [1,2]. However, it is only very recently, with the advent of fMRI and event related potentials (ERP), that the brain activity associated with face processing has been examined. A powerful methodology is required to resolve what is still one of the greatest methodological challenges in cognitive neuroscience: When dealing with complex visual stimuli, how can a brain response be attributed to a specific object category (e.g. a face), a specific feature (e.g. the eye) or a specific function? In the absence of a principled method, the specificity of response (e.g. to the face) is determined by contrast with responses from other categories (e.g. cars, furniture, hands and so forth), and informal hypotheses tested. Unfortunately, a dense correlative structure exists in the low-level visual properties of category members (e.g. luminance energy, main directions of orientation, spatial frequency composition and so forth), only a small subset of which can be controlled with a finite number of contrast categories. Consequently, the specificity of the brain response might be due to incidental input statistics, not to the category *per se*.

As a case study, consider ERP studies which reported that the upright frontal view of a human face or faces of other species, face photographs, paintings and sketches [3] elicits a negative potential 170 ms following stimulus onset (N170). The N170 is typically larger for faces than of multiple control stimuli such as humans hands, cars, birds, items of furniture [3,4]; scrambled faces, scrambled cars, butterflies [5]; non-face meaningless stimuli [6]; house [7]; cheek and back views of faces [8]; houses and hands [9,10]. In contrast, the N170 amplitude is usually smaller for upright than for inverted faces [11]. The N170 is maximally negative over posterior temporal scalp, and is probably generated by extrastriate occipitotemporal cortex regions in inferior temporal gyrus and/or the adjacent occipitotemporal sulcus [4,12]. Functionally, it has been proposed that the N170 reflects a variety of face processes, ranging from detection [4] to the categorization of emotions [13,14] and structural encoding [6,7,15,16], or perhaps more generally the configural encoding associated with expertise of novel objects and faces [16,17] or birds and dogs [18]. The N170 is less specific in 3- and 6-month-old infants [19], but its amplitude increases gradually in the early years [5,20], and is not affected by aging [21].

This body of evidence leaves little doubt that the N170 is often (but not exclusively, see [22]) associated with face processing. However, for the reasons pointed out earlier, it is still unclear exactly what face stimulus modulates the N170 amplitude. Is the stimulus the entire face or one of its features? For instance, the eyes have been found to elicit higher N170 amplitudes than the whole face and other face parts [4], and there is evidence that direction of eye gaze modulates the N170 in 4-month-old infants. However, others have found no influence of eye information on amplitude modulation [7]. This leaves unresolved the issues of which face feature(s) should be credited with the N170

response, and what status should be attributed to the N170 signature. Does the N170 reflect a diagnostic use of the face features that are most useful to resolve face categorization tasks [24], or does the N170 simply reflect an automatic response to specific face features (e.g. the eyes), irrespective of the categorization at hand?

Bubbles [24,25] is a technique that has been designed to resolve such issues of credit assignment. It works by using the stimulus (not other stimuli) as its own control for amplitude of brain response. Although we are focusing on ERPs, it is important to stress that this technique is generic and could in principle apply to other measurable brain signals. The technique only requires a parametric stimulus input space to correlate with a parametric output space. To determine the face features modulating the N170 and categorization behavior in a principled way, we applied Bubbles in two separate categorization tasks (gender and expressive or not, exnex) and compared the information determining the N170 and behavior. On each trial Bubbles samples face information (using a number Gaussian apertures) and records the behavior (correct *vs* incorrect response) and brain response (in terms of N170 amplitude) to which this sample leads. Across trials, the method can depict the face features that elicit correct *vs* incorrect behavior, and high *vs* low N170 amplitudes.

MATERIALS AND METHODS

Participants: Two Glasgow University paid observers (ML and BB) with normal or corrected to normal vision participated in the experiment.

Stimuli: Facial information was revealed through 14 2D randomly located apertures (of Gaussian shape, with sigma $= 0.22°$ of visual angle) with the constraint that each aperture remained within the area of the face (see Fig. 1 for an example of stimulus). Previous experiments revealed that 14 apertures are required to reach a minimum of 75% correct categorizations in GENDER and EXNEX [24]. Original faces were 256-gray level pictures of 10 actors (five males, five females) displaying two expressions (neutral, happy) taken under standardized conditions of illumination. Hairstyle was standardized across face pictures by replacing (using Adobe Photoshop) idiosyncratic hairstyles with a uniform unisex hairstyle, to eliminate this information.

Procedure: In one session of 4000 trials (gender) the two observers (BB and ML) determined the gender of the facial information samples (henceforth, sparse faces) by pressing the appropriate button of a two-key response box. In another 4000 trials session (exnex) they determined whether the sparse faces were expressive (happy) or not (neutral). Order of task (gender *vs* exnex) was counterbalanced across observers (BB *vs* ML). Stimuli were presented on a light grey background at the center of a computer monitor with a fixed chin rest maintaining a constant 1 m viewing distance ($4.6 \times 4.6°$ of visual angle). A trial started with the presentation of a fixation cross ($0.4°$ of visual angle) which was replaced after 500 ms by a randomly selected sparse face picture. The sparse face remained on the screen for 1500 ms and observers were instructed to respond as

quickly as possible (male *vs* female in one session, expressive or not in the other session) by depressing the appropriate response key without making mistakes. Short breaks were allowed every 100 trials. While performing the task, EEG activity was continuously recorded with sintered Ag/AgCl electrodes mounted in an electrode cap (Easy-Cap) at the scalp positions from Fz, Cz, Pz, Iz, FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, F9, F10, FT9, FT10, P9, P10, PO9, PO10, F9′, F10′, and TP9. The right mastoid (TP10) served as initial common reference, and the AFz electrode as ground. The F9′ and F10′ electrodes were positioned 2 cm anterior to F9 and F10 at the outer canthi of the left and right eye. Vertical electrooculogram (vEOG) was bipolarly registered above and below the right eye. EEG and EOG recordings were sampled at 250 Hz. Electrode impedance was kept $< 10 \, k\Omega$. All signals were recorded with a band-pass (0.05–40 Hz, $-6 \, db$ attenuation, 12 dB/octave). Analysis epochs were generated off-line, starting 200 ms prior to stimulus onset and lasting for a total duration of 800 ms. Epochs were aligned to a 200 ms pre-stimulus baseline. To sort trials, we ran artifact detection software and inspected visually each trial for ocular and nonocular artifacts. Artifact-free ERPs were low-pass filtered at 10 Hz (zero phase shift) and re-referenced to average reference, excluding the vEOG channel.

EEG analyses: We performed a single trial N170 amplitude measurement using a dipole source projection method (DSPM). We first generated a dipole model of the N170 component for each observer by averaging all artifact free trials across tasks (ERP waveforms at P9 and P10 are shown in Fig. 1b). The dipole source models were determined by using brain electromagnetic source analysis (BESA, Version 2.2.) with the 4-shell spherical head model using an 8 ms time interval around peak maximum of the N170 (for both observers from 200 to 208 ms after stimulus onset). One dipole-pair mirror-symmetrical in orientation and location was fitted in this time interval until residual variance (RV) reached a minimum (6.9% for BB and 3.7% for ML). We then projected the dipole models into the single trial data without further fitting and we calculated the mean dipole source strength for the left and right dipole across the three data points 200, 204, and 208 ms after stimulus onset. This approach is comparable with statistical component separation methods such as principal component analysis, but it takes into account biophysical aspects of the EEG signals and reduces the dimensionality of the electrode space from 32 to 2. Furthermore, the dipole projection method (DPM) eliminates EEG background noise that is not projecting in the same direction as the dipoles itself, resulting in at least a partial cleaning of the single trial data.

Behavioral analyses: Using computational analyses we then sought the face information responsible for the explicit categorization behavior and the N170 amplitude modulation. For the analysis of the behavioral data, we used Bubbles [24,25]. On each trial of a categorization task, the 14 randomly located Gaussian apertures make up a 2D mask that reveals a sparse face. Observers will tend to be correct if this information is diagnostic for the categorization task. Across trials, we computed a probability of being correct, for
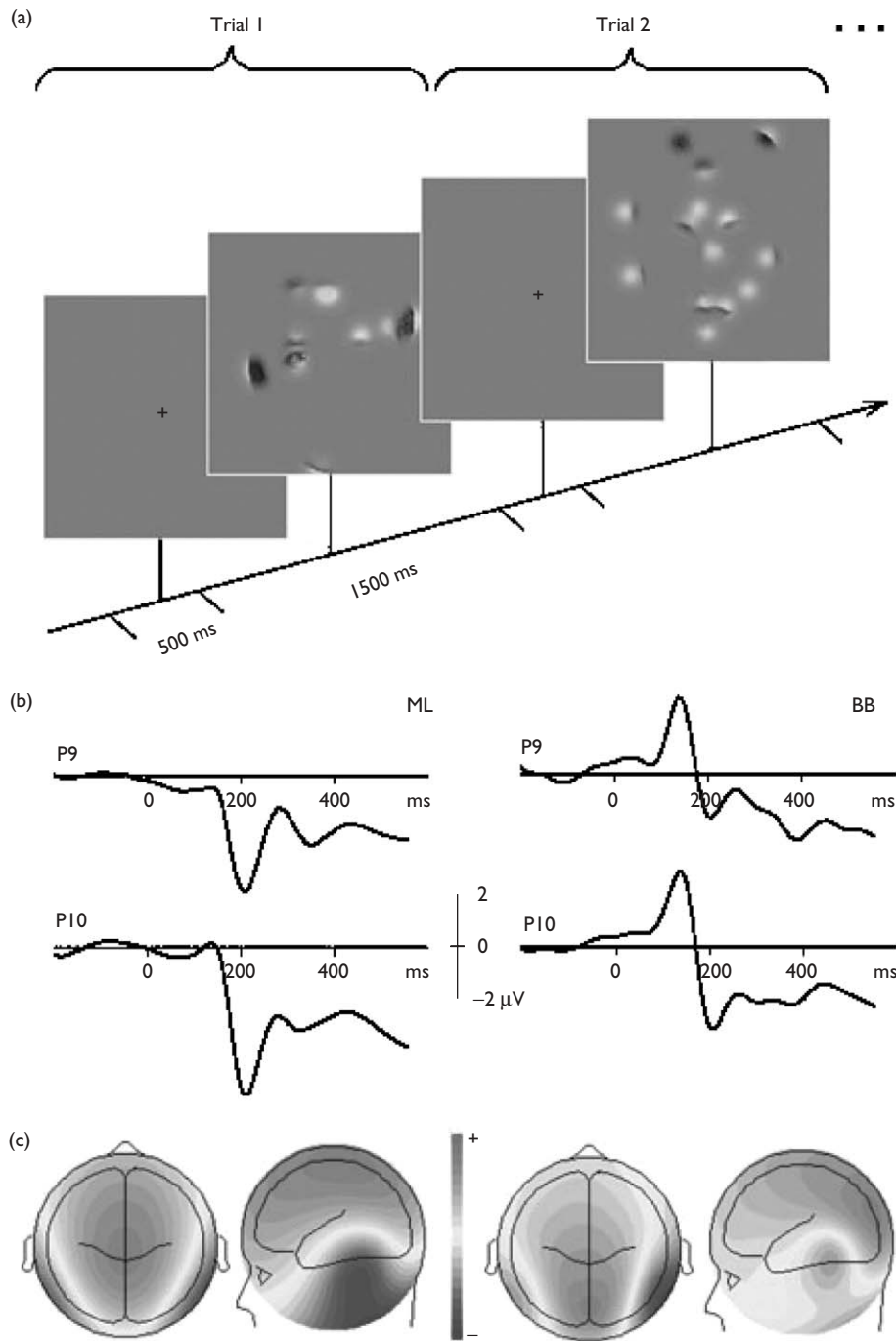
Fig. I. (a) The temporal sequence of events in each trial. (b) ERP waveforms at the P9 and PI0 electrodes for participants ML (left) and BB (right). (c) Spatial NI70 component used for single trial amplitude measurement generated from the dipole models for participants ML (left) and BB (right), the negative potentials are found in inferior-temporal regions.

each individual aperture; i.e. for each aperture, P(correct/ aperture) = frequency correct(aperture)/frequency presented(aperture). We summed the aperture masks leading to correct categorizations and divided this sum by the sum of all aperture masks (for correct and incorrect categorizations). We transformed the probabilities into Z-scores (the pictures labeled Task in Fig. 2), marked in red the

statistically significant ($p < 0.01$) probabilities, and revealed the corresponding features used to perform the gender and exnex categorizations.

To determine the face information driving the N170, we adapted Bubbles to the analysis of a distribution of N170 amplitudes. We measured the N170 in response to the sparse face in each single trial. Following the experiment,
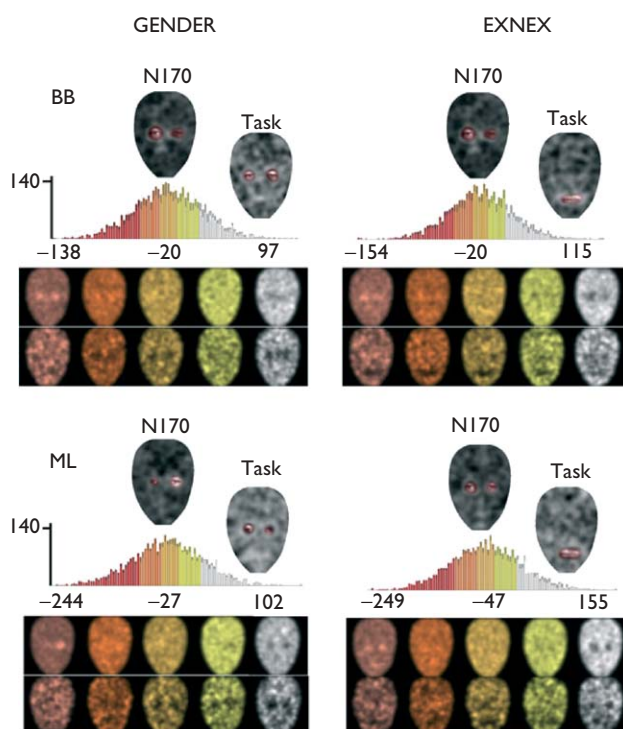
Fig. 2. Results of the experiment. The experiment was programmed with the Psychophysics Toolbox for Matlab. This figure illustrates the computation of the stimulus information eliciting NI70 and task-dependent responses for observers BB and ML resolving the gender and exnex categorizations of the same face set. Histograms represent the NI70 amplitude distributions (measured over 4000 trials) clustered into five bins (differently colored with red indicating higher amplitudes trials) of equal statistical power: the x axis indicates the NI70 amplitude, the y axis the frequency of trials. Each histogram bar is split between correct and incorrect (shaded) response trials. In color correspondence with the five ranges of NI70 amplitudes, two rows of five pictures represent the average information sample leading to each range for correct (top row) and incorrect (bottom row) trials. On top of each histogram, the NI70 picture represents in red the significant information that discriminates ($p < 0.0I$) high from low NI70 amplitudes (it is a discrimination image based on [binI + bin2]−[bin4 + bin5]). The Task picture represents in red the information used to categorize the faces ($p < 0.0I$).

we divided the N170 distribution into five bins of equal trial number (two high and two low bins symmetrically distributed around a central bin). Each bin determined an N170 amplitude interval within which we added, for each trial, the mask of apertures eliciting these amplitudes (see the distributions in Fig. 2). Following this assignment of masks of apertures to bins of amplitudes, we averaged the content of each bin to derive the average face information sample that elicited each N170 amplitude interval. The average per bin is represented in the two rows of pictures below each distribution in Fig. 2 for correct (top row) and incorrect (bottom row) categorizations, in gender and exnex, for BB and ML. To determine the information that modulates the N170, we computed a discrimination image in each task and for each observer (the pictures labeled N170 on top of each distribution). Specifically, we summed the average face information sample of the last two bins (the yellow and white low amplitude bins in Fig. 2) and subtracted this from the sum of the first two bins (the red

and orange high amplitude bins in Fig. 2); i.e. discrimination image = (bin1 + bin2)−(bin4 + bin5). For each discrimination image, we computed Z-scores, marked in red the regions of statistically significant discrimination ($p < 0.01$, in red in each image), and revealed the corresponding face features that discriminate between low and high N170 amplitudes.

RESULTS AND DISCUSSION

Summarizing the results, the analysis of behavior in gender in terms of use of information reveals that the diagnostic features were the two eyes, for both BB and ML (see Fig. 2, gender, the Task pictures). Analysis of the face information modulating the N170 also revealed that presence of eye information led to high N170 amplitudes when absence of these features led to low N170 amplitudes, for both observers. The eyes are therefore the features that discriminate between low and high N170 in gender (see Fig. 2, gender, the N170 pictures). Turning to the status of the N170 signature, the information leading to correct categorizations and high N170 amplitudes were correlated. From this correlation, one could infer that the N170 reflects the encoding of the diagnostic features (in this case, the eyes) that observers require to correctly categorize the gender of faces. This conclusion is only warranted if, using the same faces in a categorization task that requires different diagnostic face features, the N170 also responds to this other information. Analysis of behavior in the expression (exnex) task revealed that correct categorization required the diagnostic use of the mouth (see Fig. 2, exnex, the Task pictures). In contrast, the presence of information from the eyes still discriminated between small and large N170 amplitudes, when the mouth did not (see Fig. 2, exnex, the N170 pictures). Here, the information leading to correct categorizations and high N170 amplitudes was decorrelated. Thus, the N170 signature does not necessarily reflect a use of diagnostic information.

The version of the Bubbles method applied here establishes a linear mapping between information samples and responses (behavioral or neuronal, see [25] for a non-linear expansion). The advantage is that we can estimate how each sample drives performance. One could oppose that this restricts the scope of our findings because observers experience several samples of facial information, never the entire face in one sample. However, this argument neglects the fact that humans observers naturally sample visual information with saccadic eye movements. With Bubbles, information is sampled for the observer and its use is estimated for categorization. When they are compared, and this is a domain of on-going investigation, similar information picking strategies are revealed by Bubbles and eye movement studies (see [25] for discussions).

The analysis that derives the discrimination image assumes a linearity of the N170 response to face information. To confirm that this assumption was correct, we ran the distribution analysis backwards: i.e. we searched for the typical ERP signature of a face feature, for each observer, in each task. To this end, we first divided the faces into nine contiguous regions (left forehead, right forehead, left eye, right eye, nose, left cheek, right cheek, left mouth, right mouth). For each trial, we determined whether or not a

Gaussian aperture was located in each of the nine face regions. Whenever it was, we kept for this face region a record of the N170 amplitude in one of five possible amplitude bins as explained earlier. Across trials, we derived the typical N170 distribution of each face region. We found flat distributions for all face regions but the eyes. For the eyes, we observed an (approximately) linearly decreasing distribution with a maximum on the highest amplitude bin, and a minimum on the lowest amplitude bin. This was true for the two observers, in the two categorization tasks.

CONCLUSIONS

In the tasks considered here, Bubbles found that the eyes are the face features modulating the N170, even when other features drove explicit categorization behavior. With faces, to the best of our knowledge, the status of the N170 is not a response to diagnostic features, but an automatic response to the eyes [4]. We believe that the same techniques we have used to analyze the N170 could also be generalized to other ERP components (e.g. the N200 recorded intracranially over the fusiform cortex, in response to faces, to fMRI amplitude responses, or to the firing rates of individual cells, or cell assemblies. Indeed, we could gain insight into the properties of different brain responses to the same stimuli and task demands, potentially providing a dynamic picture of the mechanisms of stimulus processing.

REFERENCES

1. Farah MJ. *Behav Brain Res* **76**, 181–189 (1996).
2. Nachson I. *J Clin Exp Neuropsychol* **17**, 256–275 (1995).
3. Carmel D and Bentin S. *Cognition* **83**, 1–29 (2002).
4. Bentin S, Allison T, Puce A *et al. J Cogn Neurosci* **8**, 551–565 (1996).
5. Taylor M, McCarthy G, Saliba E and Degiovanni E. *Clin Neurophysiol* **110**, 910–915 (1999).
6. Sagiv N and Bentin S. *J Cogn Neurosci* **13**, 937–951 (2001).
7. Eimer M. *Neuroreport* **9**, 2945–2948 (1998).
8. Eimer M. *Neuroreport* **11**, 2319–2324 (2000).
9. Eimer M. *Brain Res Cogn Brain Res* **10**, 145–158 (2000).
10. Rebai M, Poiroux S, Bernard C and Lalonde R. *Int J Neurosci* **106**, 209–226 (2001).
11. Rossion B and Gauthier I. *Cogn Neurosci Rev* **1**, 63–75 (2002).
12. Schweinberger SR, Pickering EC, Jentzsch I *et al. Cogn Brain Res*, in press.
13. Campanella S, Quinet P, Bruyer R *et al. J Cogn Neurosci* **14**, 210–227 (2002).
14. De Haan M, Nelson CA, Gunnar MR and Tout KA. *Dev Neurophysiol* **14**, 495–518 (1998).
15. Rossion B *et al. Biol Psychol* **50**, 12173-12189 (1999).
16. Rossion B *et al. Neuroreport* **11**, 69-74 (2000).
17. Rossion B, Gauthier I, Goffaux V *et al. Psychol Sci* **13**, 250–257 (2002).
18. Tanaka JW and Curran T. *Psychol Sci* **12**, 43–47 (2001).
19. de Haan M, Pascalis O and Johnson MH. *J Cogn Neurosci* **14**, 199–209 (2002).
20. Taylor MJ, Itier RJ, Allison T and Edmonds GE. *Brain Res Cogn Brain Res* **10**, 333–340 (2001).
21. Chaby L, Jemel B, George N *et al. Brain Cogn* **45**, 357–377 (2001).
22. Rossion B, Curran T and Gauthier I. *Cognition* **85**, 189–196 (2002).
23. Schyns PG. *Cognition* **67**, 147–179 (1998).
24. Gosselin F and Schyns PG. *Vis Res* **41**, 2261–2271 (2001).
25. Schyns PG, Bonnar L and Gosselin F. *Psych Sci* **13**, 402–409 (2002).