Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 2010–2020
Vision Research
www.elsevier.com/locate/visres

Mutual information of image fragments predicts categorization in humans: Electrophysiological and behavioral evidence ☆

Assaf Harel [a,*], Shimon Ullman [b], Boris Epshtein [b], Shlomo Bentin [a,c]

[a] Department of Psychology, Hebrew University of Jerusalem, Israel
[b] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel
[c] Center of Neural Computation, Hebrew University of Jerusalem, Israel

Received 2 January 2007; received in revised form 28 March 2007

Abstract

Computational models suggest that features of intermediate complexity (IC) play a central role in object categorization [Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*, 682–687.]. The critical aspect of these features is the amount of mutual information (MI) they deliver. We examined the relation between MI, human categorization and an electrophysiological response to IC features. Categorization performance correlated with MI level as well as with the amplitude of a posterior temporal potential, peaking around 270 ms. Hence, an objective MI measure predicts human object categorization performance and its underlying neural activity. These results demonstrate that informative IC features serve as categorization features in human vision.

Keywords: Categorization; Object recognition; Features; Human performance; Event-related potentials (ERPs)

1. Introduction

Object category exemplars may differ substantially in visual aspects such as shape, size and texture. For instance, in the category of cats there are many different breeds, each with its own distinctive visual characteristics. Despite of this striking variability, humans are able to easily generalize across the various exemplars defining a "cat" category, and to distinguish this category from other categories such as "dog".

Multiple areas in the visual cortex contribute to visual object categorization. Following initial processing in the primary visual cortex (V1), the visual processing leading to object categorization engages occipito–temporal structures as well as structures in the ventral temporal lobe. Together, these structures form the ventral pathway of the visual system (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). Although the activity of these structures is highly interactive, different stages of processing are reflected in their differential sensitivity to different features of the object (Tanaka, 1996). Primitive features such as edges and bars at different orientations and spatial frequency scales are extracted and represented in the primary visual cortex (V1) (De Valois, Albrecht, & Thorell, 1982; Hubel & Wiesel, 1968). Further downstream, cells in areas V4/TEO respond selectively to moderately complex features (Tanaka, 1996, 2003) and even further downstream, in the ventral temporal lobe, cells respond to complete views representing the integrated shapes of the objects (Grill-Spector & Malach, 2004; Kreiman et al., 2006).

Ample knowledge has accumulated over the years regarding the functional tuning characteristics of V1, on the one hand, and the ventral temporal lobe on the other. In contrast, relatively little is known about the nature of

☆ This study was funded by NIMH Grant R01 MH 64458 to S.B. S.U. and B.E. were supported by ISF Grant 7-0369 and EU IST Grant FP6-2005-015803. We thank Anni Levitt for skillful research assistance.
* Corresponding author. Fax: +972 2 582 5659.
E-mail address: assafusa@mscc.huji.ac.il (A. Harel).

0042-6989/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2007.04.004

intermediate representations and about the functional characteristics of the neural systems involved with processing these representations. It has been suggested that the intermediate representations along the ventral visual pathway are important for basic-level object categorization (Logothetis & Sheinberg, 1996; Tjan, 2001) but the nature of these intermediate features remained unclear. Several types of visual features have been proposed in the past for encoding objects and categories, ranging from simple local image patterns such as wavelets, Gabor filters, edges and blobs (Mel, 1997; Riesenhuber & Poggio, 1999; Wiskott, Fellous, Krüger, & von der Malsburg, 1997) to abstract three-dimensional shape primitives, such as Geons (Biederman, 1987). A common aspect of most previous features is that they were generic in nature, that is, a fixed, small set of features types was used to represent all objects and categories.

Alternatively, a recent approach proposed that visual object categorization is based on representing shapes by a combination of category-specific shared sub-structures called fragments (Ullman, Vidal-Naquet, & Sali, 2002). According to this approach the fragments are extracted from image examples by maximizing the amount of information they deliver for categorization. This information can be formally expressed by the equation $I(C,F) = H(C) - H(C|F)$, where $I(C,F)$ denotes the mutual information (MI) (Cover & Thomas, 1991) between the fragment F and the category C of images, and H denotes entropy. C and F in this scheme are binary variables: F denotes whether a certain feature is found in the image and C denotes whether the image belongs to the target category C. Thus, the usefulness of a fragment for representing a category is measured by the reduction in uncertainty about the presence of an object category C in an image by the possible presence of that fragment F in the image. This value is evaluated for a large number of candidate fragments, and the most informative ones are selected. Notably, in contrast to many previous approaches, which suggest either small local (Mel, 1997; Wiskott et al., 1997) or global features (Turk & Pentland, 1991), the optimal features for different categorization tasks according to this model are typically of intermediate complexity (IC) including intermediate size at high resolution and larger size at intermediate resolution (for more details see Section 2).

According to this informative fragment-based model, the level of MI contained in an image patch (a fragment) conveys its usefulness for categorization. If, at least in an approximate form, MI maximization is used as a neuronal coding principle underlying intermediate stages of object categorization, then the model predicts that features with high measured MI will cause higher neuronal activation and better categorization performance in humans than features with low MI. Because the MI of a visual feature is determined by simple similarity between the fragment and the image, this direct relation is not predicted by other approaches (see Section 4).

The goal of the present study was to assess empirically the influence of the MI contained in IC fragments on human performance and neural activity manifested during categorization. For this purpose, we first extracted fragments with different levels of MI from a number of object categories. We then recorded event-related potentials (ERPs) elicited by IC fragments of different categories and varying MI levels while participants categorized these fragments as parts of faces or cars. Given their excellent time resolution, ERPs can disclose the time course of the neural events involved in object categorization. Previous ERP studies of object visual categorization suggested that categorical distinctions can be found in ERPs as early as about 100 ms (e.g. Thorpe, Fize, & Marlot, 1996). However, these studies investigated the categorization of complete objects rather than intermediate, incomplete representations. Notably however, ERP studies of face processing showed that the face-sensitive N170 component is elicited by isolated face features (Bentin, Allison, Puce, Perez, & McCarthy, 1996) regardless of their configuration (Zion-Golumbic & Bentin, in press). On the basis of these studies we expected that the influence of MI on the neural mechanisms involved in the categorization of IC features will be evident during the first 200 ms of stimulus. Specifically, we predicted that the amplitude of a categorically distinctive negative component analogous to the N1/N170 would be increased as a function of the feature's MI.

2. Methods

2.1. Participants

A total of 48 volunteers participated in the three experiments for course credit or monetary reward. All participants were Hebrew University students with normal or corrected to normal visual acuity and no history of psychiatric or neurological disorders. Participants signed an informed written consent according to the institutional review board of Hebrew University. In Experiment 1 (the explicit categorization behavioral experiment) participants were 20 students (11 females), aged 20–34. In Experiment 2 (the explicit categorization ERP experiment), participants were 14 students (11 females), aged 19–23, and in Experiment 3 (the implicit categorization ERP experiment) participants were 14 students (7 females), aged 20–32.

2.2. Stimuli selection procedures

Informative fragments were extracted from training images using the algorithm described by Ullman et al. (2002), and briefly summarized below. Fragments were extracted from a total of 1000 face images, 350 car images, and 2000 non-class images downloaded from the web, with image-sizes ranging from 150×200 to 200×250 pixels. The non-class images were a general collection from different classes of objects selected at random. The fragment-selection process initially extracts a large number of candidate fragments at multiple positions, sizes and scale from the class images. The information supplied by each candidate fragment is estimated by detecting it in the training images. To detect a given fragment F in an image, the fragment is searched by correlating it with the image. If the normalized cross-correlation value at any location exceeds a certain threshold θ, then F has been detected in the image ($F = 1$), otherwise, $F = 0$. A binary variable $C(I)$ is used to represent the class, namely, $C(I) = 1$ if the image I contains a class example, and 0 otherwise. For each

candidate fragment, the amount of information it delivers about the class is then estimated based on its detection frequency within and outside the class examples supplied to the algorithm. The delivered information is a function of the detection threshold, the threshold for each fragment is therefore adjusted individually to maximize the delivered information $I(F;C)$. Finally, a subset of the most informative fragments is selected. To avoid redundancy between similar features, fragments are selected successively, where at each stage, the fragment that contributes the largest amount of additional information is added to the set of selected fragments. This selection process was found in theoretical and practical comparisons to be highly effective for selection of features from a large pool of candidates (Fleuret, 2004). In selecting low-information fragments for the testing, the fragments used in this study were maximal in the sense that they did not contain smaller sub-regions with higher mutual information than the full fragment.

The most informative fragments found for different categorization tasks are typically of intermediate complexity, including intermediate size at high resolution and larger size at intermediate resolution. The reason is that, to be informative, a feature should be present with high likelihood in class examples, and low likelihood in non-class examples. These requirements are optimized by IC features: a large and complex object fragment is unlikely to be present in non-class images, but its detection likelihood in novel class examples also decreases; conversely, simple local fragments are often found in both class and non-class images.

The stimuli used in all three experiments were 500 car fragments, 500 face fragments and 500 non-class fragments (see Fig. 1 for examples of stimuli). For face fragments the MI range was 0.05–0.65 while for car fragments it was 0.04–0.35. This difference probably reflects intrinsic differences between the face and car categories, most notably their within-class variability. Therefore, to allow direct comparison, we divided the continuous MI range of each category of fragments into five consecutive discrete levels of equal size, ranging from 1 (lowest MI level) to 5 (highest MI level) with 100 different fragments within each level. This ordinal (rather than continuous) scale allowed direct comparison between categories. In addition we also assessed the influence of the continuous MI values on performance in correlational analyses, calculated separately for each category. For the remaining of the text, we refer to the discrete variable as MI level, and to the continuous values as MI proper.

For each category, the gray level histograms, the average number of edges (computed by a common edge detector, Canny, 1986), and the number of 'interest points' (computed by the Harris detector, Schmid, Mohr, & Bauckhage, 2000) were compared for fragments with MI below and above the respective category MI mean. In all categories the low- and high-MI sets were not statistically different.

The non-class fragments were included in all experiments in order to establish a baseline activity associated with the categorization of image fragments from various undefined categories of objects. However, since, by definition, non-class fragments did not represent a homogeneous category, they were not included in the main analysis. Rather, the categorization of non-class fragments was analyzed separately (see Supplementary material). Finally, 150 horse fragments were added as targets in the implicit categorization ERP experiment but not analyzed (see Section 2.4). The horse fragments were selected using the same computational procedures.

2.3. Explicit categorization task and procedure

Participants were asked to decide whether a fragment was a part of a face, a part of a car, or a part of some other object by pressing one of three pre-designated buttons. The 1500 stimuli were presented sequentially in a fully randomized order in 10 blocks of 150 trials each with a short (up to a minute) break between blocks. At the end of each block a feedback of the participant's accuracy of performance was provided. Each stimulus was presented for 100 ms and then a response was required. The next trial commenced only after a response was made. Stimuli were presented at fixation and seen from a distance of approximately 60 cm. The behavioral and the ERP experiments were identical in all the above details and differed only in that the ERP experiment was run in a sound attenuated and electrically isolated booth whereas the behavioral experiment was run in a room of similar settings except for electrical shielding.

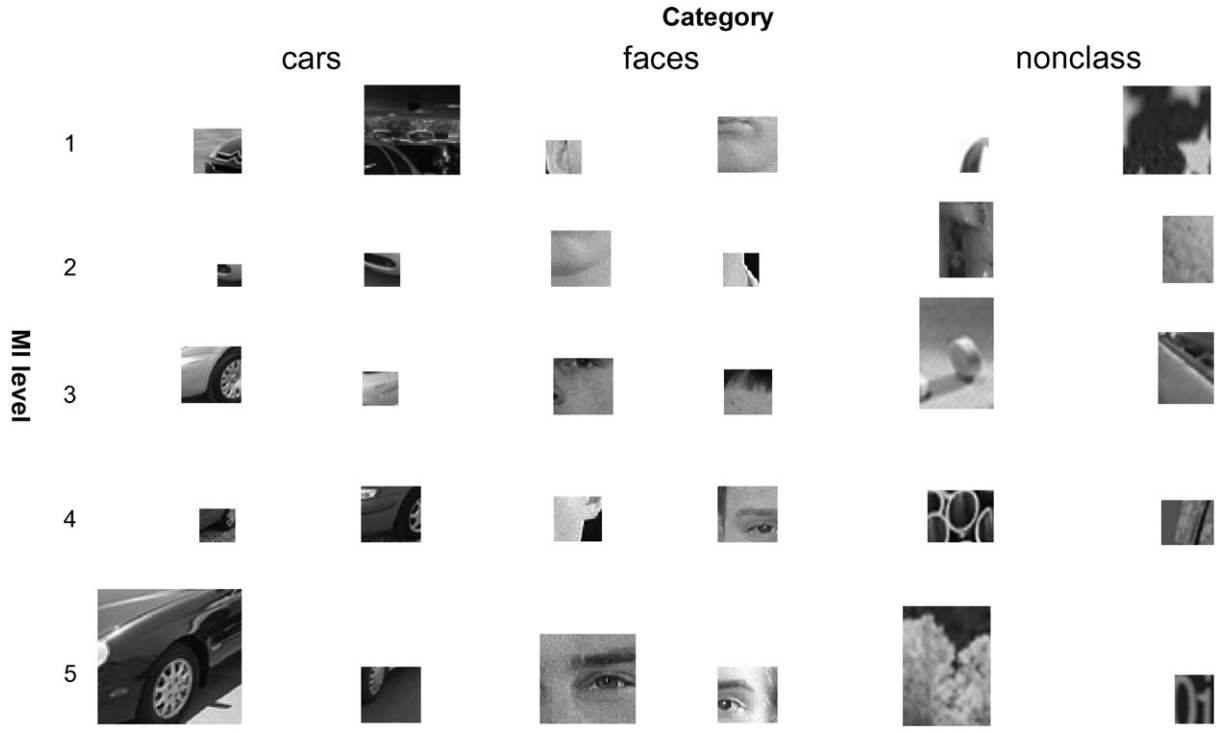

Fig. 1. Examples of intermediate complexity fragments used in the three experiments. Fragments from the car, face, and nonclass categories were used in all experiments and horse fragment (not shown here) were used only in the implicit categorization ERP experiment. Fragments in each category are ordered according to their level of MI, ranging from 1 (lowest MI level) to 5 (highest MI level).

2.4. Implicit categorization task and procedure

The task was oddball target monitoring in which car, face, nonclass and horse fragments were presented one after another and participants were requested to press a button each time a part of a horse appeared on the screen. This procedure ensured that all stimulus categories of interest in this study were equally task-relevant, that is, they were all distracters to be ignored. The stimuli were the same stimuli used in the explicit categorization experiments with the addition of the horse fragment targets (10% of total number of stimuli). The 1650 stimuli were presented sequentially in a fully randomized order in 10 blocks of 165 trials each with a short (up to a minute) break between blocks. At the end of each block, feedback was provided to the participants about their accuracy. Stimuli were presented for 100 ms, with 1500 ms ISI and were presented at fixation and seen from a distance of approximately 60 cm.

2.5. EEG recording

The EEG analog signals were recorded continuously by 64 Ag–AgCl pin-type active electrodes mounted on an elastic cap (ECI) according to the extended 10–20 system (American Electroencephalographic Society, 1994), and from two additional electrodes placed at the right and left mastoids, all reference-free. Eye movements, as well as blinks, were monitored using bipolar horizontal and vertical EOG derivations via two pairs of electrodes, one pair attached to the external canthi, and the other to the infraorbital and supraorbital regions of the right eye. Both EEG an EOG were sampled at 1000 Hz using a Biosemi Active II digital 24-bits amplification system with an active input range of −262 mV to +262 mV per bit without any filter at input. The digitized EEG was saved and processed off-line.

2.6. ERP data processing and analysis

Raw data was 1.0 Hz high-pass filtered (24 dB) and referenced to the tip of the nose. Eye movements were corrected using an ICA procedure (Jung et al., 2000). Remaining artifacts exceeding ±100 μV in amplitude or containing a change of over 100 μV in a period of 50 ms were rejected. Artifact free data was then segmented into epochs ranging from 250 ms before to 800 ms after stimulus onset for all conditions. ERPs resulted from averaging the segmented trials separately in each condition. The averaged waveforms were smoothed by applying a low-pass filter of 17 Hz (24 dB) and baseline-corrected based on the time between 150 and 50 ms before stimulus onset.

For each subject the peaks of the P1 and N270 were determined (based on the filtered waveform) as the most positive peak between 80–150 ms and the most negative peak between 200 and 320 ms, respectively. Subsequent visual scrutiny ensured that the highest voltage values represented real peaks rather than end points of the epoch. Based on scrutiny of the present N270 distribution, the statistical analysis was restricted to posterior–lateral regions. The amplitudes and latencies of the N270 at sites P8, PO8 and P10 within each hemisphere yielded the dependent variables for ANOVA. The characteristic scalp distribution of the N270 in each condition was estimated by spherical spline interpolations (interpolation order = 4). ANOVAs with repeated measures were separately applied on P1 and N270 amplitudes and latencies. The factors were Category (car fragments, face fragments), MI (levels 1–5), Hemisphere (right, left) and site (P7/8, PO7/8 and P9/10). For factors with more than two levels, *p*-values were corrected for non-sphericity using the Greenhouse–Geisser correction (for simplicity, the uncorrected degrees of freedom are presented).

3. Results

3.1. Experiment 1: Explicit categorization—behavioral experiment

Since all previous explorations of Ullman et al.'s (2002) model were based on computer simulations, a first experiment was designed to explore its psychological reality in humans.

As presented in Fig. 2b, the RT decreased as a function of MI level ($F_{4,76} = 52.83$, $P < .0001$), and there was no main effect of category ($F_{1,19} = 1.95$, $P = .18$). However, a significant MI by category interaction ($F_{4,76} = 7.17$, $P < .0001$) revealed different RT curves for the two object categories. Whereas for face fragments the RTs decreased monotonically with increasing levels of MI, for car fragments they decreased in a more step-like function; the RTs were similar for the two lower levels of MI, both higher than the RTs to the three higher MI levels, which did not differ among themselves. Post-hoc comparisons showed that for face fragments all differences between two successive MI levels were significant ($P < .05$) while for car fragments the only significant difference ($P < .05$) between successive MI levels was between the second and third levels.

As a more continuous approach to the relation between MI and RT we calculated the mean RT of correct responses across all participants for each fragment within each category. These individual fragments' RTs were correlated with their absolute MI value. The correlation analyses revealed highly significant negative correlations in the well-defined categories ($r = -.42$; $P < .001$, and $r = -.30$; $P < .001$ for face and car fragments, respectively) and no correlation for non-class fragments ($r = .02$; NS).

Categorization accuracy increased monotonically with successive levels of MI ($F_{4,76} = 115.17$, $P < .0001$) and a main effect of category ($F_{4,76} = 2.64$, $P < .05$) revealed an advantage in categorization accuracy of face fragments over car fragments across all MI levels (Fig. 2a). An MI by category interaction was marginally significant ($F_{4,76} = 2.64$, $P = .06$).

3.2. Experiment 2: Explicit categorization—ERP experiment

The neurophysiological correlate of the influence of MI on the fragments' categorization was assessed by comparing the ERPs elicited by fragments of different MI levels. An identical explicit categorization task was conducted using the same stimuli and design as in Experiment 1. In addition to ERPs, categorization accuracy and RTs were also collected.

The earliest and most conspicuous manifestation of the MI influence on ERPs was the modulation of the amplitude of a negative potential peaking at about 270 ms (N270; Fig. 3a). This component was distributed bilaterally at posterior–temporal sites (Fig. 3b). Its amplitude was modulated by MI ($F_{4,52} = 14.85$, $P < .0001$) and category ($F_{1,13} = 14.32$, $P < .01$) with the N270 increasing with MI level, and larger for face than for car fragments. The MI effect on the N270 amplitude was qualified by its interaction with category ($F_{4,52} = 48.72$, $P < .015$). Post-hoc univariate contrasts between successive levels of MI within each category were conducted (Fig. 3c). For face

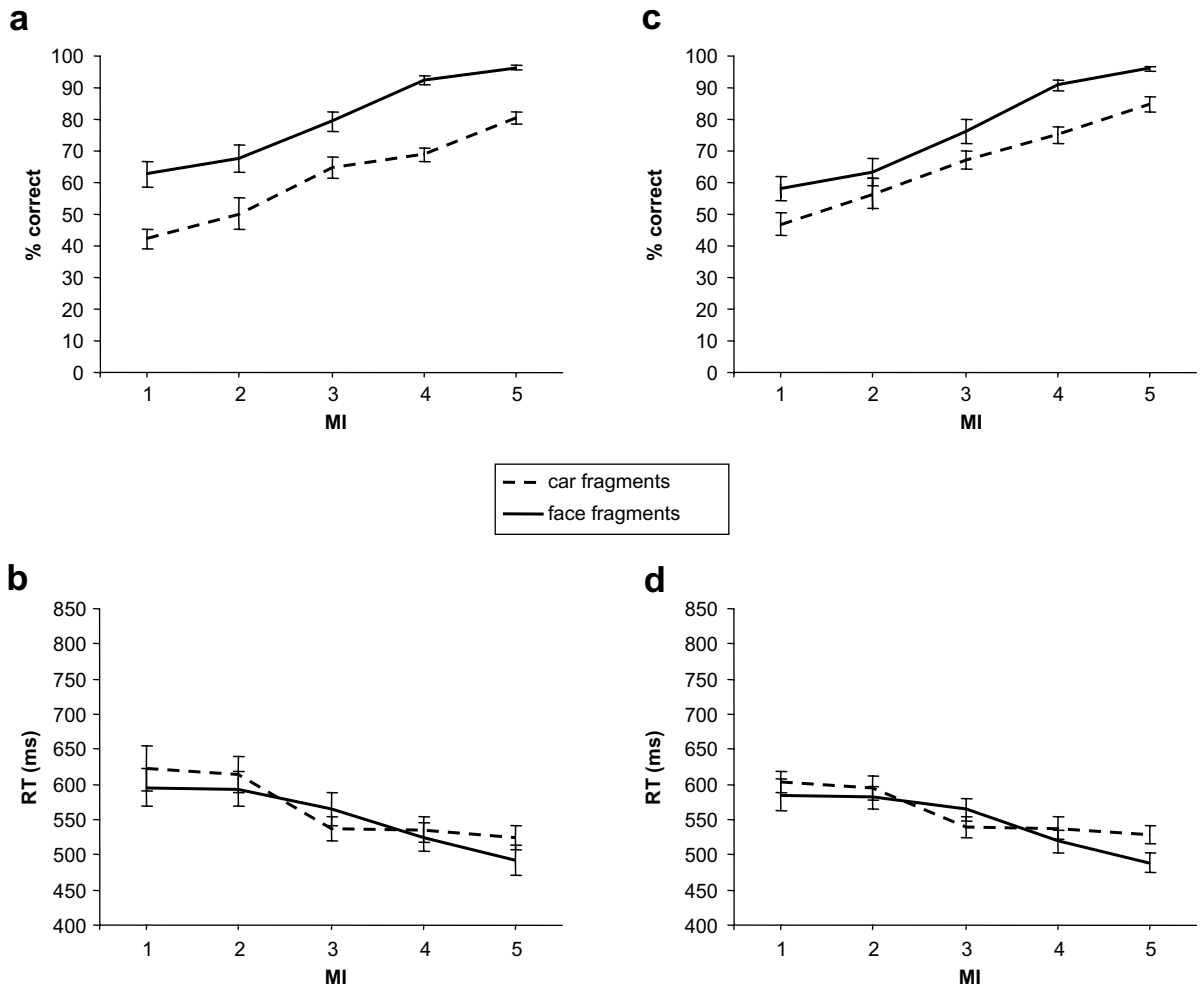

Fig. 2. Mean reaction times and accuracy rates of categorization of car and face fragments (dashed and full lines, respectively) as a function of MI level in the explicit categorization behavioral experiment (a and b) and in the explicit categorization ERP experiment (c and d). MI levels are in ascending order, 1 representing the lowest level, 5 representing the highest level. Error bars indicate *SEM*. Note the high similarity between the shapes of the RT and accuracy curves in the two experiments.

fragments, starting at level 3, the N270 amplitude increased continuously with the MI level; significant differences were found between third and forth MI levels ($F_{1,13} = 9.36$, $P = .009$) and between forth and fifth levels of MI ($F_{1,13} = 8.54$, $P = .01$). For car fragments, the amplitude increased in a step-like function, separating the MI levels into low (MI levels 1–2) and high (MI levels 3–5), as significant differences were found between second and third MI levels ($F_{1,13} = 8.64$, $P = .01$) but not within these clusters. Similar analyses of the N270 latency showed no significant interaction effects.

To rule out the possibility that the current N270 effect reflects putative systematic differences in low-level properties of the stimuli, we analyzed the effect of MI on the earlier positive peak (P1) elicited by the fragments. P1 is an early ERP component with a source in the extrastriate visual cortex and sensitive to the amount of sensory stimulation provided by a stimulus (Gonzales, Clark, Fan, Luck, & Hillyard, 1994; Hillyard & Picton, 1987). In contrast to the N270, there was no significant P1 amplitude change between any successive MI levels ($F_{4,52} < 1.00$) and there was no interaction ($F_{4,52} < 1.00$). This finding corroborates the hypothesis that the effect of MI on N270 was not related to possible differences in physical stimulus attributes that might have correlated with the MI level.

The pattern of MI effects on performance during the ERP experiment was consistent with the electrophysiological results and replicated the results of Experiment 1. The accuracy and speed of categorization of both car and face fragments increased as function of MI ($F_{4,52} = 122.24$, $P < .0001$ and $F_{4,52} = 30.30$, $P < .001$, respectively). As in the ERP results, for RT there was an interaction of category and MI ($F_{4,52} = 3.85$, $P < .05$) reflecting the difference between the RT curves for the two object categories. Whereas for face fragments the RTs decreased monotonically with increasing levels of MI, for car fragments they decreased, again, in a more step-like function. Like in Experiment 1 within the car-fragments category the RTs were similar for the two lower levels of MI, both higher than the RTs to the three higher

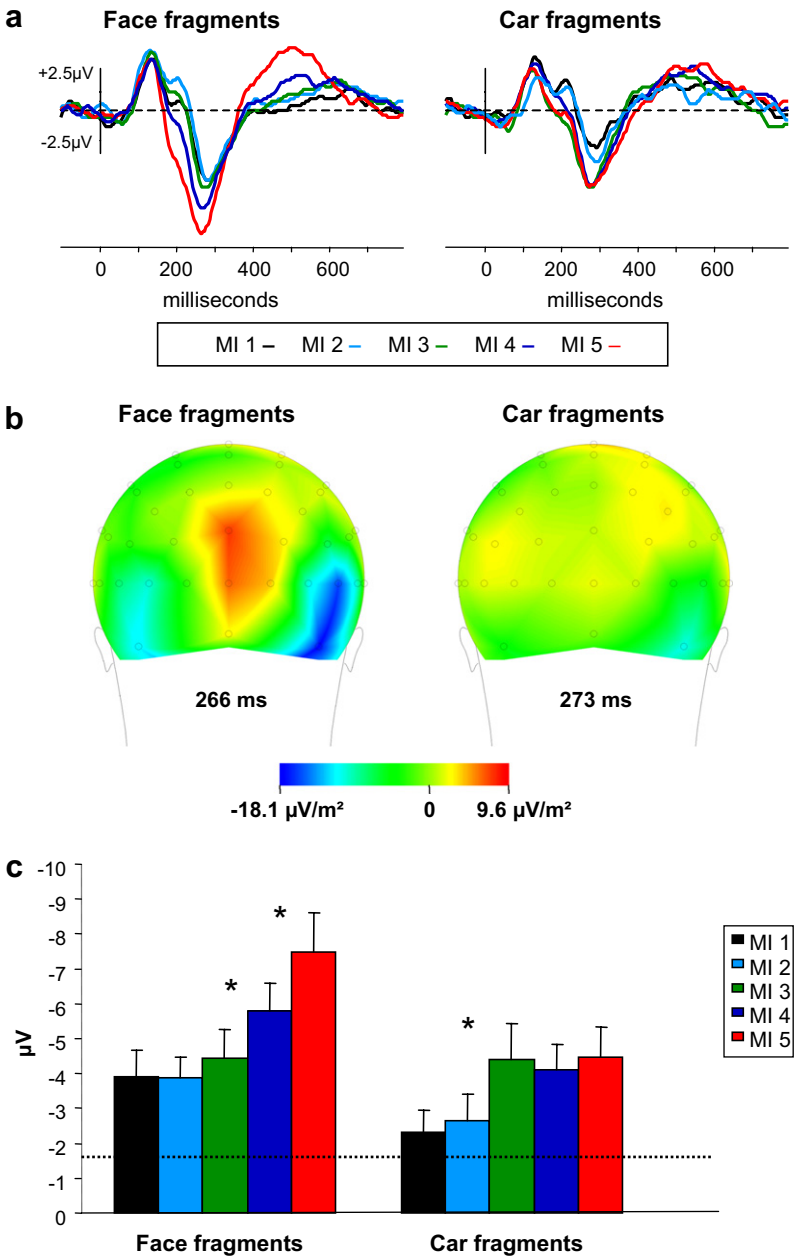

Fig. 3. The modulation of the N270 waveform by MI level and category at posterior lateral scalp sites in the explicit categorization ERP experiment. (a) Group averaged ERPs ($n = 14$) for the five MI levels for car and face fragments at the right hemisphere (data is plotted for the P10 electrode). (b) Scalp current source density (CSD) topographical maps of the N270 for car and face fragments in the highest level of MI at peak latencies recorded at P10 electrode. As can be seen, the N270 effect is most prevalent in lateral occipito–temporal electrodes (P7, P9, PO7, P8, P10 and PO8) and is more pronounced for face fragments than for car fragments. (c) Mean N270 peak amplitudes of the five MI levels for car and face fragments (error bars indicate *SEM*). Peak amplitudes were averaged across hemisphere and electrode as none of these factors interacted with category or MI level. Significant differences ($P < .05$) between successive levels of MI within each category are denoted by asterisk (see text for details).The dashed line represents N270 mean peak amplitude for the nonclass object fragments (mean −1.67 μV, *SEM* ±0.15); see Supplementary material.

MI levels, which did not differ among themselves (Fig. 2d). A significant main effect of category on accuracy revealed that across MI, categorization was more accurate for face fragments than for car fragments ($F_{1,13} = 41.24$, $P < .0001$; Fig. 2c).

The within-category correlations between the absolute MI value of each fragment and its mean categorization time (for correct responses only) also replicated Experiment 1. The correlations were negative and highly significant for both well-defined categories ($r = -.37$; $P < .001$, and $r = -.28$; $P < .001$ for face and car fragments, respectively) while close to zero for non-class fragments ($r = -.06$; NS).

3.3. Experiment 3: Implicit categorization ERP experiment

Together, the behavioral and the electrophysiological data suggest that MI is a reliable diagnostic measure of

object categorization that predicts the efficiency of categorization performance and the level of neural activity associated with it. Moreover, the relative late latency of the MI level effect on N270 and the lack of P1 modulation suggest that the mutual information between a feature and a category affects categorization rather than low-level perception. However, it is not entirely clear what kind of categorization mechanisms are reflected by the N270 and its modulation by MI. On the one hand, it could reflect a task-induced strategy whereby facing decision uncertainty, participants utilize the MI to assign a fragment to one of the pre-designated categories. On the other hand, object categorization could be a task-independent default of visual perception (cf. Smith, Bentin, & Spalek, 2001), in which case, the N270 modulation might reflect the use of MI in this process.

To distinguish between the two accounts above, we conducted a second ERP experiment using an oddball target-detection paradigm in which the participants were not required to explicitly distinguish between faces and cars. The face, car and nonclass fragments used in the previous ERP experiment were intermixed with fragments of a new category, horses. Horses were the only pre-designated target category and the participants were requested to press a button each time a horse fragment was identified. Ten percent of the fragments were "horse" while non-target fragments occurred in the rest of the trials with equal probability. The instructions did not specify the different categories of the non-targets and thus, although car, face and non-class fragments were explicitly distinguished from horse fragments the differentiation among these non-target categories was not necessary for performing the task, hence implicit.

Supporting the "categorization by default" account, the analysis of the ERPs elicited by the non-target fragments revealed a pattern that was very similar to the explicit categorization ERP experiment (Fig. 4a).

Although the N270 was slightly delayed relative to the explicit categorization case (peaking at 285 ms for faces and 292 ms for cars), its amplitude was modulated by MI as in the previous experiment, was similarly distributed across the scalp (Fig. 4b), and was identically modulated by MI. The main effect of MI level was significant ($F_{4,52} = 13.25$, $P < .001$) as was the effect of category (amplitudes higher for face than car fragments, $F_{1,13} = 20.90$, $P < .01$). Again, the pattern of the MI effect on the N270 amplitude was different for faces and cars ($F_{4,52} = 13.25$, $P = .05$). Post-hoc univariate contrasts between successive levels of MI within each category were conducted (Fig. 4c). For face fragments, the N270 amplitude increased gradually with the MI level; significant differences were found between second and third MI levels ($F_{1,13} = 4.60$, $P = .05$) and between forth and fifth level of MI ($F_{1,13} = 8.54$, $P = .01$). For car fragments, the amplitude again increased in a step-like function, separating the five MI levels into low (MI levels 1–2) and high (MI levels 3–5). Significant differences were found only between second and third MI levels ($F_{1,13} = 12.58$, $P = .004$). An analysis of the occipital P1 showed a main effect of category ($F_{1,13} = 8.79$, $P = .01$) and of MI level ($F_{4,52} = 6.73$, $P = .02$). However, there was no MI by category interaction ($F_{4,52} < 1.00$) and post-hoc univariate contrasts between successive levels of MI across categories showed no significant differences ($P > .05$) between any two successive MI levels. In summary, N270 responses increased in monotonic and predicted manner with MI level during fragments' processing although there was no need for explicit further assignment of non-target fragments to specific categories. This outcome supports the notion that the enhanced neural activation associated with MI utilization underlies a default perceptual categorization process that is independent of task-related strategies.

4. Discussion

ERP and behavioral data in the present experiments indicate that mutual information between feature and category is predictive of the neural activity associated with perceptual categorization, as well as of human categorization performance. Although the features selected by the algorithm were simply fragmented image patches rather than clearly delineated whole parts of objects (as would be a wheel of a car, or a human eye, for example), they were sufficiently diagnostic for object categorization by humans. The diagnosticity of the features was reflected in performance, and was correlated with the neural activity that they elicited, as reflected by a posterior negative ERP component (N270); the amplitude of the N270 was modulated by the MI level of the fragments. For face fragments the N270 amplitude increased continuously with successively increasing levels of MI. For car fragments the amplitude increased in a step-like function distinguishing between the two low and the three high-MI levels. The modulation of the N270 amplitude by MI level mirrored performance indicating that the pattern of the relation between MI and performance has neural origins.

As noted in the introduction, categorical distinctions were found in the visual system as early as 100–150 from stimulus onset (Thorpe et al., 1996; Van Rullen & Thorpe, 2001). Accordingly, we expected that the utilization of the MI for basic-level categorization should occur somewhere during the first 200 ms of stimulus processing. However, although the onset of the MI-sensitive, category selective waveform peak was earlier than 200 ms,[1] its peak amplitude of was around 270 ms post-stimulus onset while no conspicuous negative component (N1) was found prior to this potential. Specifically, it is intriguing that fragments of faces, although categorized correctly by the participants, did not elicit the N170 potential normally found in response to faces (Bentin et al., 1996; George, Evans, Fiori,

[1] Corresponding, indeed, with Thorpe et al., 1996 who looked for the earliest time in which there was a divergence between 'class' and 'non-class' analyzing component's onset rather than their peaks.

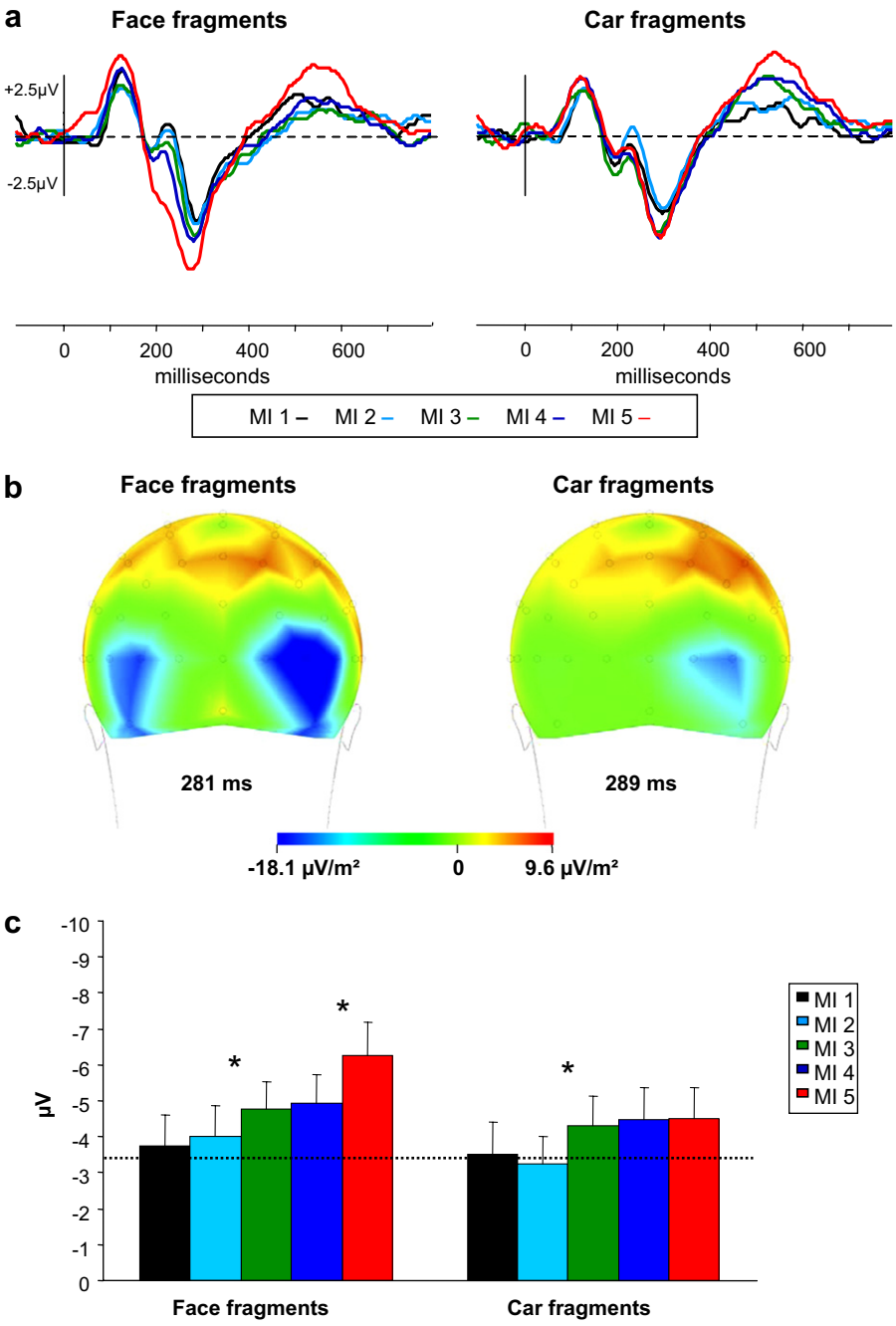

Fig. 4. Modulation of the N270 waveform by MI level and category at posterior lateral scalp sites in the implicit categorization ERP experiment. (a) Group averaged ERPs ($n = 14$) for the five MI levels for car and face fragments at the right hemisphere (data is plotted for the P10 electrode). (b) Scalp current source density (CSD) topographical maps of the N270 for car and face fragments in the highest level of MI at peak latencies recorded at P10 electrode. As can be seen, the N270 effect is distributed exactly as when the face-car categorization is explicit. (c) Mean N270 peak amplitudes of the five MI levels for car and face fragments (error bars indicate SEM). Peak amplitudes were averaged across hemisphere and electrode as none of these factors interacted with category or MI level. Significant differences ($P < .05$) between successive levels of MI within each category are denoted by asterisk. The dashed line represents N270 mean peak amplitude for the nonclass object fragments (mean −3.57 μV, *SEM* ±0.22).

Davidoff, & Renault, 1996). The absence of this effect is particularly intriguing because robust N170 effects were reported not only in response to complete faces, but also in response to face components presented in isolation (Bentin et al., 1996; Smith, Gosselin, & Schyns, 2004; Zion-Golumbic & Bentin, in press). A possible account for this unexpected pattern is that the N270 is, in fact, a delayed N170. This account is supported by the higher amplitude for face than car fragments across MI levels, by the posterior lateral distribution of the N270, (which is similar to that of the N170; Figs. 3b and 4b), and by the fact that the N170 to isolated face components is also somewhat delayed compared to full faces (Bentin et al., 1996). According to this account, the delay in latency may reflect

additional processing needed to categorize isolated features when the activation of the face processing mechanism is lower than normal. This could happen because in the present study the category exemplars were represented by isolated fragments. In contrast, in previous studies in which ERP distinctions between categories were found earlier, the stimuli were full-shaped category exemplars, which according to our model are in effect combinations of informative fragments. It is possible, that when whole objects (or a configuration of fragments) are seen, bottom-up utilization of the mutual information of the IC features becomes faster, either by facilitation between features or by recurrent activity generated by higher levels in the visual processing hierarchy (Rao & Ballard, 1999). This would explain the discrepancy between our findings and the fast object categorization found in previous studies.

Alternatively, it is possible that the N270 is not a delayed N170 but an independent ERP component that reflects a general categorization principle of MI extraction from visual objects in the environment. The lack of an N170 effect could be due to the fact that the face fragments were insufficient for the pre-determined tuning characteristics of the face perception system, and thus did not elicit face-sensitive neural mechanisms. Still, it is conceivable that extensive experience with a particular category, such as faces, enables higher sensitivity to changes in MI levels. Thus, the categorical differences in N270 amplitudes elicited by face and car fragments could reflect different levels of visual experience with various objects (Palmeri & Gauthier, 2004). Further research using fragments of other object categories of varying experience levels is needed to clarify the issue of the effect of experience and learning on the utilization of MI for categorization.

Importantly, the modulation of N270 by MI level is unlikely to reflect low-level visual differences between the different levels of MI. First, the latency of the N270 was probably too late to be directly modulated by physical stimulus dimensions such as size, luminance, contrast etc. Second, the earlier P1 component which is usually sensitive to such dimensions did not vary across MI levels. Similar arguments were used to account for the amplitude modulation of another occipito–temporal ERP component peaking at 290 ms, which increased gradually with the degree of perceptual closure of line-drawn objects (Doniger et al., 2000). Third, as reveled by a separate analysis, the MI level had no influence on the N270 amplitude elicited by non-class fragments that were equivalent with the car and face fragments in low-level visual properties (see Supplementary material). Finally, the comparison of image statistics revealed that low and high-MI fragments did not differ on low-level visual properties (see Section 2 for details).

The previous arguments speak against the possibility that the N270 amplitude has been modulated in this study by low level stimulus properties. Another argument to be considered is that the modulation of the N270 reflects a general effect of task difficulty rather than being specific to MI. This argument is weakened by reports showing that, in contrast to the present pattern, the amplitude of the N2b, a negative component peaking during the same time range, is increased by difficulty in visual discrimination tasks (Senkowski & Hermann, 2002). This trend is opposite to the current findings that show a positive correspondence between the N270 amplitude and categorization accuracy.

Excluding low level visual factors, on the one hand, and general difficulty effects, on the other hand, leads to the conclusion that a high-level perceptual categorization mechanism accounts for the present findings. Since the instructions in Experiments 1–2 emphasized the distinction between two basic-level categories (‘‘faces’’ and ‘‘cars’’) we believe that in these experiments MI influenced basic-level categorization. Note, however, that each of these basic-level categories might also be conceptualized as superordinate categories (e.g. ‘‘human’’ and ‘‘vehicle’’ or even ‘‘living’’ and ‘‘non-living’’; see Mandler, Bauer, & McDonough, 1991). While the obvious option for superordinate categorization exists there are reasons to reject it. First, there is considerable evidence suggesting that the basic-level is the preferred level of categorization (e.g. Archambault, Gosselin, & Schyns, 2000; Johnson & Mervis, 1997; Jolicoeur, Gluck, & Kosslyn, 1984; Rosch, Mervis, Gray, Johnson, & Boyes-Bream, 1976). Second, computationally, the present fragments were extracted and their MI rated using training at the basic-level, that is, by contrasting a particular class with general non-class images. Still, claims about the level of categorization addressed in this study should be considered with caution, and the question of how MI relates to different hierarchical levels of categorization remains a question for future studies.

The correspondence between MI level and categorization is not trivial. We found that an *objective* measure, the mutual information between a feature (fragment) and a class predicts neural activity (as reflected by the N270) as well as categorization performance. This relation is consistent with common sense, since an image region that is objectively informative can also be expected to be useful in human vision. It is not clear, however, that fragment MI, a simple measure used to assess the amount of information delivered by an image fragment, would be sufficient for predicting neural activation and categorization performance. The MI computation was based on using the fragment directly as a feature, based on its frequency within and outside the class of interest. In different theories of recognition, such as geon-based (Biederman, 1987), eigenfaces (Turk & Pentland, 1991), internal transformations (Shepard & Metzler, 1971; Tarr, 1995) and others, such information measure would not be predictive of either neural activity or performance because no measure of informativeness is assigned to the basic elements of the image. The patch-information measure used here allowed us to directly compare, for example, two different parts of a car and predict their usefulness for categorization, both at the neural and at the behavioral level.

As revealed in the implicit categorization experiment, the utilization of MI for categorization, as indexed by the N270 modulation, is at least in part a task-independent process. Although in that experiment the task did not require the distinction between non-target categories, the data indicated that the visual system was still sensitive to the information content of the face and car fragments. Moreover, the similar patterns of modulation of the N270 in the explicit and in the implicit categorization tasks suggest that MI utilization is the default of the visual system. Note however, that the ~20 ms delay of the N270 peak in the implicit categorization experiment might indicate that task-related explicit strategies can influence the efficiency of using the MI. At any rate, the similar pattern of results suggests that despite its late latency the N270 reflects primarily sensory rather than post-sensory, target-decision processes (Johnson & Olshausen, 2005).

In conclusion, the present results demonstrate that a simple objective measure of mutual information between a visual feature and a category predicts categorization performance by humans and its underlying neuronal activity as measured by N270 ERP component. This outcome supports the notion that features of intermediate complexity are the basis for swift basic level categorization in humans.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2007.04.004.

References

American Electroencephalographic Society. (1994). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 11, 111–113.

Archambault, A., Gosselin, F & Schyns, P.G. (2000). A natural bias for the basic-level? *Proceedings of the XXII Meeting of the Cognitive Science Society* (pp. 60–65). Lawrence Erlbaum: Hilldsale, NJ.

Bentin, S., Allison, T., Puce, A., Perez, A., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*, 551–565.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Canny, J. A. (1986). Computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*, 679–698.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*, 545–559.

Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E., et al. (2000). Activation timecourse of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure processes. *Journal of Cognitive Neuroscience, 12*, 615–621.

Fleuret, F. (2004). Fast binary feature selection with conditional mutual information. *Journal of Machine Learning Research, 5*, 1531–1555.

George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal and moderately scrambled faces. *Brain Research Cognitive Brain Research, 4*, 65–76.

Gonzales, C. M. G., Clark, V. P., Fan, S., Luck, S. J., & Hillyard, S. A. (1994). Sources of attention-sensitive visual event-related potentials. *Brain Topography, 7*, 41–51.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*, 20–25.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience, 27*, 649–677.

Hillyard, S. A., & Picton, T. W. (1987). Electrophysiology of cognition. In F. Plum (Ed.), *Handbook of physiology: Section 1. The nervous system* (pp. 519–584). Bethesda, MD: Waverly Press.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126*, 248–277.

Johnson, J. S., & Olshausen, B. A. (2005). The earliest signatures of object recognition in a cued-target task are postsensory. *Journal of Vision, 5*, 299–312.

Jolicoeur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 19*, 31–53.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology, 111*, 1745–1758.

Kreiman, G., Hung, C. P., Kraskov, A., Quiroga, R. Q., Poggio, T., & DiCarlo, J. J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron, 49*, 433–445.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19*, 577–621.

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology, 23*, 263–298.

Mel, B. W. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation, 9*, 777–804.

Palmeri, T., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5*, 291–304.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*, 79–87.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Bream, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Schmid, C., Mohr, R., & Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision, 37*, 151–172.

Senkowski, D., & Hermann, C. S. (2002). Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. *Clinical Neurophysiology, 113*, 1742–1753.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Smith, M. C., Bentin, S., & Spalek, T. M. (2001). Is spreading of semantic activation truly automatic? The influence of task and response delay. *Journal of Experimental Psychology: Learning Memory and Cognition, 27*, 1289–1298.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2004). Receptive fields for flexible face categorizations. *Psychological Science, 15*, 753–761.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109–139.

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex, 13*, 90–99.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three dimensional objects. *Psychonomic Bulletin & Review, 2*, 55–82.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Tjan, B. S. (2001). Adaptive object representation with hierarchically-distributed memory sites. *Advances in Neural Information Processing Systems, 13*, 66–72.

Turk, M. A., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3*, 71–86.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*, 682–687.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge: MIT Press.

Van Rullen, R., & Thorpe, S. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13*, 454–461.

Wiskott, L., Fellous, J. M., Krüger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions of Pattern Analysis and Machine Intelligence, 19*, 775–779.

Zion-Golumbic, E. & Bentin, S. (in press). Dissociated neural mechanisms for face detection and configural encoding: Evidence from N170 and Gamma-band oscillation effects. *Cerebral Cortex*. Available online, October 24, 2006.