Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms

HALINE E. SCHENDAN,[a,b] GIORGIO GANIS,[c] AND MARTA KUTAS[a,b,c]

[a]Program in Neurosciences, University of California at San Diego, La Jolla, USA
[b]Interdisciplinary Program in Cognitive Science, University of California at San Diego, La Jolla, USA
[c]Department of Cognitive Science, University of California at San Diego, La Jolla, USA

Abstract

The nature and early time course of the initial processing differences between visually matched linguistic and nonlinguistic images were studied with event-related potentials (ERPs). The first effect began at 90 ms when ERPs to written words diverged from other objects, including faces. By 125 ms, ERPs to words and faces were more positive than those to other objects, effects identified with the P150. The amplitude and scalp distribution of P150s to words and faces were similar. The P150 seemed to be elicited selectively by images resembling any well-learned category of visual patterns. We propose that (a) visual perceptual categorization based on long-term experience begins by 125 ms, (b) P150 amplitude varies with the cumulative experience people have discriminating among instances of specific categories of visual objects (e.g., words, faces), and (c) the P150 is a scalp reflection of letterstring and face intracranial ERPs in posterior fusiform gyrus.

Descriptors: Event-related potentials, Visual perceptual categorization, Word perception, Face perception, Perceptual skill learning, Posterior fusiform gyrus

Performance differences between linguistic and nonlinguistic visual objects have been demonstrated in behavioral studies using a variety of tasks (e.g., Potter & Faulconer, 1975); these findings suggest processing differences. Neuropsychological research on people with focal brain lesions and studies using event-related potentials (ERPs) support this inference. For example, visual agnosia and prosopagnosia, impairments of object and face recognition, respectively, have been observed without alexia, an impairment of written word recognition (e.g., Farah, 1994). Early intracranial ERPs, peaking as early as 150 ms, are selectively elicited by words but not faces or other nonlinguistic visual objects in the posterior fusiform gyrus (Allison, McCarthy, Nobre, Puce, & Belger, 1994). Scalp ERPs at later times have demonstrated effects that reflect higher level processing differences. For example, patterns of scalp N200 effects are distinct to words and faces during analysis of emotional information (Deldin, 1996; Miller, 1996), and scalp N400 congruity effects are distinct to words and objects during sentence comprehension (Ganis, Kutas, & Sereno, 1996).

In the studies reported herein, we focused on the nature and onset of the first perceptual processing differences between linguistic and nonlinguistic visual images that reflect more than differences in simple, low-level image properties, such as spatial frequency spectra or aspect ratio. Such an effect may indicate the earliest neural processing stage that is language-specific. In the first experiment, we recorded ERPs while people viewed images that were wordlike (words, nonwords, and pseudofont strings), objectlike (objects and pseudo objects), or both (icon strings). These different types of images were matched for several low-level visual properties, such as contrast and size, and one higher order perceptual grouping property, namely closure (Koffka, 1935, p. 151).

The first ERP effect that appeared to reflect differences in higher order image properties began around 125 ms and peaked around 155 ms; wordlike images elicited much more positivity than objectlike ones. This positivity and these effects were maximal at Cz and apparent at all electrodes except occipital ones. We identify these early effects with the main ERP peak with which they overlap at least initially (i.e., a P150).

The P150 thus seems selective for images resembling linguistic visual patterns. As such, the P150 seems to indicate an early perceptual process that differentiates between linguistic and nonlinguistic stimuli. However, a similar difference has been reported for faces versus other objects (Bötzel & Grüsser, 1989; Jeffreys, 1989). Thus, the P150 may not be language-specific.

Rather, we hypothesized that the P150 may index the earliest point during visual processing when neural operations are activated selectively by images resembling well-learned categories of visual patterns, of which words and faces are premier examples. Consistent with this hypothesis, the overall pattern of responsivity and latency range of the scalp P150 resembles that reported for intracranial potentials recorded from the human posterior fusiform gyrus. Specifically, whereas in one section of the posterior fusiform gyrus these intracranial ERPs are much larger to words than faces or other nonlinguistic objects, in adjacent yet distinct sections of the posterior fusiform gyrus, they are much larger to faces than other objects or words (Allison, McCarthy, et al., 1994).

Recently, Jeffreys (1996) pointed out that the relationship between the scalp P150 and these intracranial potentials to faces needs to be established experimentally. One supporting piece of evidence would be a P150 to words, as we observed. However, showing that P150s to words and faces have similar scalp distributions would provide stronger evidence. Accordingly, to test the hypothesis that word and face scalp P150s reflect primarily the activity of anatomically and functionally similar neural operations, we compared the P150s evoked by visually matched word and face images directly for the first time. We found that the stimulus selectivity and scalp topography, latency, and polarity of word and face scalp P150s are consistent with the idea that they reflect similar perceptual neural processes probably in posterior fusiform gyrus.

METHODS

Experiment 1

Materials

Six image types were presented (Figure 1A) on a computer monitor centered in front of each participant. Images of "objects" were line drawings of familiar objects (Snodgrass & Vanderwart, 1980). "Pseudoobjects" were constructed from the parts of each of these objects. All character strings, namely wordlike and icon strings, were matched for character frequencies and length. "Words" were highly imageable, concrete nouns. "Nonwords" were constructed by selecting randomly from the set of letters in all the "words" with the constraint that each random letterstring contains at least one illegal trigram (Kucera & Francis, 1967). "Pseudofont" strings were constructed from 26 characters visually matching the word font. "Icon strings" were constructed from 26 highly simplified line drawings of objects (i.e., icons) excluding ones in the object image condition. Characters in the pseudofont and icon strings had the same spatial dimensions and string positions as in words and nonwords. Moreover, the sets of pseudofont and icon strings each had the same relative character frequencies and string lengths as in words and nonwords.

All image types were matched further on several visual dimensions: (a) brightness, which was moderate; (b) contrast, which was maximum; (c) hue, all appeared in green against a black background; (d) area, they subtended an area that was approximately 9° square; (e) line thickness, namely spatial frequency amplitude spectra collapsed across all orientations (De Valois & De Valois, 1988); and (f) closure (Koffka, 1935, p. 151), a higher order perceptual grouping property, that is, the characters composing the wordlike and icon strings and the objectlike images all formed closed figures.

Procedure

For each image type, participants saw 80 instances requiring no overt response (nontargets) and 8 requiring a button press by the dominant hand (targets) whenever the image was (a) drawn in red, (b) immediately repeated, or (c) the name or the depiction of a fruit or vegetable. These tasks were performed in three separate blocks. For counterbalancing, the images were divided into three sets; each
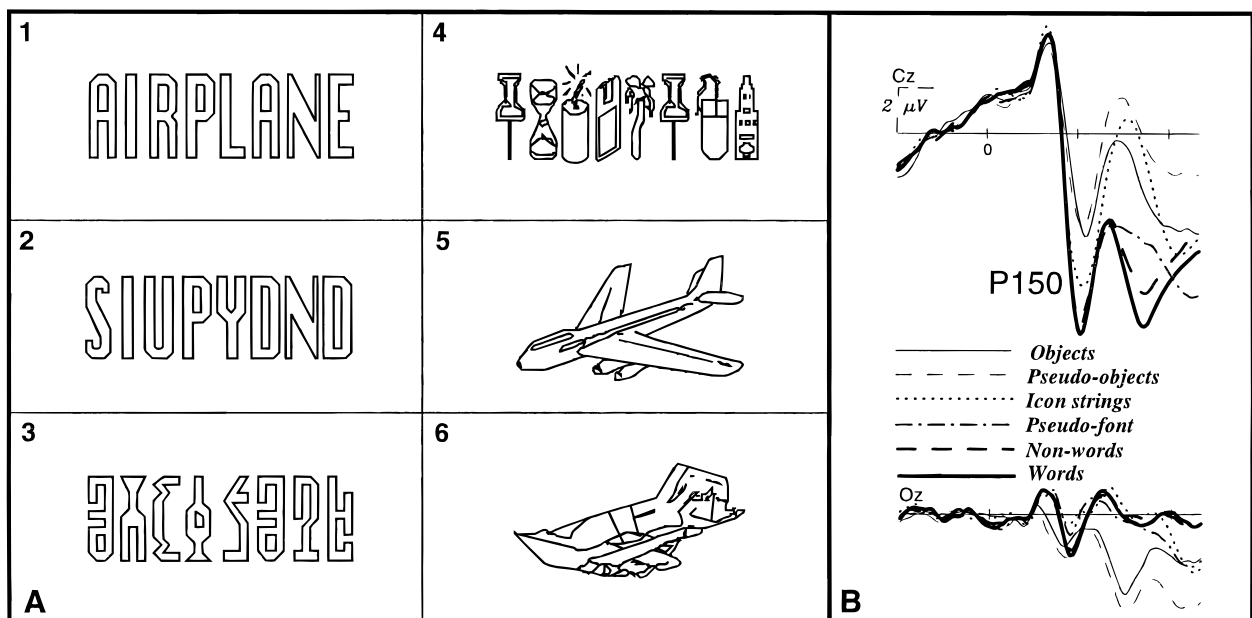
Figure 1. (**A**) Examples of the six image types shown in Experiment 1: (1) words, (2) nonwords, (3) pseudofont strings, (4) icon strings, (5) objects, or (6) pseudoobjects. These are also examples of six of the seven image types shown in Experiment 2. (**B**) Grand-average ERPs from −150 ms to 350 ms recorded in Experiment 1. P150 amplitude, which is maximal at Cz, differentiates wordlike (or letter stringlike) images from nonlinguistic objectlike images.

participant saw all three sets, but the task performed on each set and the order of the tasks were counterbalanced across individuals. Each image was preceded by a 1,200-ms fixation period that consisted of a + sign appearing for 400 ms followed by a blank screen for 800 ms. Each image was then presented for 800 ms, after which the screen went blank for 2,600–3,000 ms in between trials.

Participants

Twelve UCSD undergraduates (English monolinguals, right-handed, 18–28 years) volunteered for cash or course credit.

Experiment 2

Materials

All six image types from Experiment 1 (Figure 1A) plus line drawings of faces (Figure 2, 80 instances/type) were presented that had been matched for the low-level visual and perceptual properties described for Experiment 1. However, in this experiment, all the pictures were presented as black figures against a light-medium gray background. The faces were in frontal views and half were female. Most of the faces wore a neutral (56%) or mildly positive emotional (27%) expression.

Procedure

For each image type, participants saw 80 nontarget instances and 8 targets. Unlike Experiment 1, the target and nontarget images did not systematically differ. Rather, the infrequent target images were indicated only by a prompt ("??LIKE??" shown in red pixels for 3,000 ms) that appeared 1,000–1,750 ms after the target image disappeared. When they saw the prompt, participants rated how much they liked the immediately preceding image on a 4-point scale, pressing one of four buttons with the designated finger of the dominant hand. Participants were instructed to perform this task on all images passively, responding overtly only to images followed by the prompt. Each image was preceded by a 1,300–1,700-ms fixation period that consisted of a + sign appearing for 700 ms (in red pixels) followed by a blank screen for 600–1,000 ms. Each image was then presented for 700 ms, after which the screen went blank for 1,000–1,750 ms in between trials.

Participants

Nineteen UCSD undergraduates (native English speakers, 17 right-handed, 18–26 years) volunteered for cash or course credit.

Electrophysiological Recordings and Analyses

Electrophysiological Recording

Electroencephalographic (EEG) activity was sampled at 250 Hz (bandpass = 0.01–100 Hz) from 26 tin electrodes in a geodesic
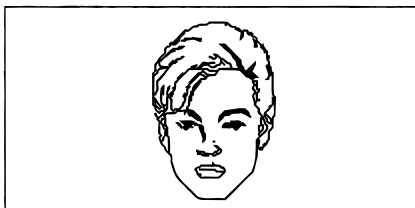

Figure 2. For Experiment 2, along with the six image types from Experiment 1 (see Figure 1A), there was a seventh image type: faces.

arrangement across a plastic cap and an electrode over the right mastoid. To monitor eyeblinks, an electrode on the upper cheek bone was centered underneath the right eye. All these electrodes were referenced to a left mastoid electrode. To monitor eye movements, bilateral electrodes on the outer canthi of both eyes were referenced to each other.

ERP Averaging

For each participant, the ERPs were (a) constructed off-line by averaging the EEG to nontarget images, excluding trials with horizontal eye (4%) or other movement artifacts (18%, including eyeblinks); (b) time locked to image onset with a 150-ms prestimulus baseline; (c) re-referenced to the average of left and right mastoids; and (d) in Experiment 1 only, low-pass filtered at 50 Hz.

Raw ERP Analyses

Onset latency. In both experiments, to determine the onset latency of image type effects, we analyzed the mean amplitude of ERPs within 11, successive, nonoverlapping, 12-ms time windows between 0 and 132 ms. There were several image type comparisons: (a) objectlike images, that is, objects versus pseudoobjects; (b) wordlike images, that is, words versus nonwords versus pseudofont strings; (c) words versus icon strings; (d) objects versus icon strings; (e) words versus objects; and, for Experiment 2 only, also (f) words versus faces; and (g) faces versus objects.

P150 modulation by image type. To better determine the nature of the P150 effect, the mean amplitude of the ERPs to each image type was also analyzed between 125 and 175 ms in both experiments, and, to focus on the P150 peak in Experiment 2, mean amplitude was also analyzed across a 4-ms epoch between 152 and 156 ms, which is an interval spanning two fixed time points delimited by the peak latencies of P150s to faces (152 ms) and words (156 ms) at the Cz maximum in the grand average. In overall analyses of variance (ANOVAs), all six (Experiment 1) or seven image types were compared (Experiment 2). In addition, the five (Experiment 1) or seven (Experiment 2) image-type comparisons were also conducted, as described for the onset latency analyses.

For Experiment 2, to evaluate a possible occipital polarity reversal of the P150, namely an N150, positive or negative peak latency between 125 and 175 ms was also compared between adjacent lateral electrode pairs (i.e., N150 at OL, OR vs. P150 at temporal sites 4, 8 or parietal sites, 14, 17). This analysis was conducted separately for ERPs to words, faces, and objects.

Aspect ratio and string length. To evaluate the effect of the dimension of image size in Experiment 1, two other ERP analyses over early time windows were conducted with images further subdivided. One way to describe image size is to calculate the *aspect ratio*, that is, the image height divided by the width. This analysis was conducted only on ERPs to objectlike images (namely, objects and pseudoobjects), and the three ratios compared were (a) aspect ratio ≤ 0.7, (b) $0.7 <$ aspect ratio < 1.2, and (c) aspect ratio ≥ 1.2. Another way to assess image size is to analyze the effect of string length on wordlike and icon strings. The three lengths compared had five, six, or seven to nine characters.

Difference Wave Analyses

Onset latency. For Experiment 2, to evaluate further when the effects observed at the P150 peak begin, we conducted analyses on the shape of the scalp distribution of normalized difference ERPs (i.e., McCarthy & Wood, 1985). The pattern of modulation at each

electrode across the head, namely the shape of the scalp distribution, is closely related to the configuration of its underlying neural generators. Specifically, if the shape of two scalp distributions differs, then the configurations of the underlying neural generators must differ (Nunez, 1981; Srebro, 1996). Note that the converse does not necessarily hold because distinct source configurations can produce scalp distributions with the same shape (e.g., Amir, 1994).

The standard method to compare the shape of scalp fields entails performing an ANOVA on the normalized amplitude of difference ERPs. Differences in the shape of the scalp distributions are then captured by the interaction between electrode and condition (McCarthy & Wood, 1985), which for the analyses herein refers to time window. Amplitude normalization is necessary because identical neural configurations, if activated to different degrees, can nonetheless produce significant interactions between electrode and time window (McCarthy & Wood, 1985). The crucial point is that the shape of the scalp distribution of a difference ERP (i.e., irrespective of its absolute size) is the signature of the set of neural generators that are differentially sensitive to, for example, words and objects.

For this analysis, the shape of the scalp distribution of the difference ERP was calculated for faces minus objects, words minus objects, and words minus faces. We compared these measures at the P150 peak (152–156 ms) versus eight nonoverlapping 4-ms time windows from 120 ms to 152 ms.

Temporally overlapping later ERPs. For Experiment 2, to evaluate whether temporally overlapping later slow waves, such as N200, P300, N400, or P600, might be responsible for the "P150" modulations, we compared measures taken at the P150 peak (152–156 ms) versus successively later times after 175 ms (the end of the P150 mean amplitude epoch of 125–175 ms) until 650 ms. The mapping of time windows to componentry was as follows: 152–156 ms (P150 peak) and, for slow ERPs, 175–250 ms (N200, e.g., Miller, 1996), 250–350 ms (P300, e.g., Rohrbaugh, Donchin, & Eriksen, 1974), 350–450 ms (N400, e.g., Kutas & Hillyard, 1980), and 450–550 and 550–650 ms (P600, e.g., Stuss, Picton, Cerri, Leech, & Stethem, 1992).

As for the other difference ERP analyses, the measure used was the shape of the scalp distribution (Nunez, 1981; Srebro, 1996) of the normalized difference ERP (McCarthy & Wood, 1985). This measure was calculated for faces minus objects, words minus objects, and words minus faces. The basic idea is that, if the same processes underlie effects at the times compared, an ANOVA comparing the shape of the scalp distribution of normalized difference waves at two times should reveal no reliable interaction between time window and electrode.

The reasoning for this set of analyses is as follows. For simplicity, we will take as an example the difference ERPs for faces minus objects and the relationship between the P150 and the N400; the same reasoning applies to the other analyses. Assume that the P150 is not modulated by our experimental manipulation. In other words, assume that the P150 is the same for faces and objects and that the apparent modulation of the P150 is entirely due to a modulation of, for example, an incoming N400. At any particular time, the shape of the scalp distribution can be captured by calculating the normalized difference in ERP amplitude to faces minus objects at all electrodes. This measure can be calculated at the P150 peak (152–156 ms) and during the N400 (350–450 ms). Assume that the difference between faces and objects observed at the P150 peak is due entirely to an incoming N400. If this is true,

then there should be no change in the shape of the scalp distributions of the normalized difference ERP between these two times. In other words, the interaction between electrode and time window should not be significant.

ANOVAs

General. Greenhouse–Geisser adjustments to degrees of freedom were applied to ANOVAs to correct for violation of the assumption of sphericity, $\alpha = 0.05$. To adjust the alpha level for multiple planned comparisons, the familywise error rate was divided by the number of comparisons, unless otherwise noted.

Raw ERPs. For each analysis, one ANOVA included midline (Fz, Cz, 23, Oz) and another ANOVA included lateral (all other) electrodes to assess hemispheric asymmetry. Within-individual factors were image type, task, electrode, and, for lateral electrode ANOVAs, hemisphere.

For planned comparisons between image types, data were collapsed across task if there were no significant task effects in the ANOVAs with all image types compared. For image type contrasts on lateral or midline ANOVAs, using a family error rate of 0.05, the error rate per contrast was 0.01 (Experiment 1) or 0.0071 (Experiment 2). For image type contrasts on each pair of electrodes (conducted to localize effects), using a family error rate of 0.05, a sharper "step-up" Bonferroni procedure determined accepted p values (Hochberg, 1988); these values are specified in the footnotes for each set of electrode contrasts. For onset latency analyses, using a family error rate of 0.1, the error rate per 8-ms time window contrast was 0.009.

Difference waves. For the ANOVAs on the amplitude of normalized difference ERPs, all analyses compared the P150 peak (152–156 ms) with another time window, and the factor of main interest was time window rather than image type. Also, for these analyses, all 26 lateral and midline electrodes were included in a single ANOVA to maximize power. However, for the analyses comparing the P150 peak with later ERPs, additional ANOVAs were performed that included only 23 electrodes (the occipital sites OL, OR, and Oz were excluded to examine only the P150 rather than also the N150). Because this analysis included fewer electrodes, significant interactions between image type and electrode are less likely to be found; thus, when found, significant interactions provide even stronger evidence that there are real distribution differences between the P150 and later components.

For onset latency analyses, using a family error rate of 0.1, the error rate per 4-ms time window contrast was 0.0125, and the error rate per time window contrast between the P150 peak and later ERPs was 0.02.

RESULTS

Experiment 1

Onset Latency

Words versus objects. The comparison between words and objects is most directly relevant for evaluating processing differences between linguistic and nonlinguistic images. We found that the ERPs to words and objects first diverged between 100 and 120 ms at midline posterior sites modulating the N1(00). The focal nature of this effect was indicated by significant interactions of image type and electrode along the midline, $df(3, 33)$, between 96 and 108 ms (lateral $F < 2.1$; midline $F = 7.05$, $p < .006$, $\epsilon = 0.5397$)

and 108 and 120 ms (lateral $F < 2$; midline $F = 7.55$, $p < .004$, $\epsilon = 0.5944$); the main effects were not reliable between 96 and 108 ms (lateral $F < 1$; midline $F < 1$) and 108 and 120 ms (lateral $F < 1$; midline $F < 3$). Contrasts on midline electrode pairs indicated that this effect was posterior; between 96 and 108 ms, there was a significant main effect of image type at posterior but not at anterior midline sites, 23,Oz $F = 7.38$; Fpz,Cz $F < 1$; between 108 and 120 ms, there were no significant main effects, 23,Oz $F < 2$; Fpz,Cz $F < 1$, but the interaction of image type with electrode was significant at posterior but not at anterior midline sites, 23,Oz $F = 14$, $p < .004$; Fpz,Cz $F < 1$.[1] This effect and the other early posterior midline N1(00) effects probably reflected minor differences in simple image properties, such as spatial frequency spectra.

More important, it was not until around 125 ms that the broad positivity referred to as the P150 began to differentiate between words and objects. This difference was indicated by significant main effects of image type between 120 and 132 ms, lateral $F(1,11) = 16.37$, $p < .002$; midline $F(1,11) = 12.63$, $p < .005$. This effect seemed to be widespread because there were no significant interactions of image type with electrode at this time (lateral $F < 3$; midline $F < 7.4$). However, pairwise electrode contrasts revealed that the main effects of image type were not yet reliable at occipital sites where the P150 was minimal (at Oz $F < 1$) or absent (where an N150 appeared instead at OL,OR $F < 1$), whereas the main effects were highly reliable at more anterior sites where the P150 was largest (e.g., Fpz,Cz $F = 33.01$, $p < .001$).

Other image type comparisons. There were no significant ERP differences between any other image types until around 90 ms. A small difference then appeared between objects and pseudoobjects at lateral occipital sites modulating the P1(00), as indicated by a marginally significant lateral interaction of image type and electrode between 84 and 96 ms, $F(10,110) = 4.71$, $p = .0076$, $\epsilon = 0.2992$. Among wordlike images, in contrast, there were no early differences between words, nonwords, and pseudofont strings (i.e., at least until 175 ms; see also P150 analyses). ERPs to wordlike images and icon strings were not significantly different until after 125 ms; however, between 108 and 132 ms, there was a tendency for icon strings to diverge from the wordlike images, a small Oz N1(00) modulation, that was indicated by almost significant interactions of image type and electrode between words and icon strings along the midline between 108 and 120 ms, $F = 7.84$, $p = .0128$, $\epsilon = 0.4983$, and 120 and 132 ms, $F = 8.70$, $p = .0100$, $\epsilon = 0.4548$, $df(3,33)$. Also within this time, ERPs to icon strings and objects diverged; specifically, a small Oz N1(00) difference was indicated by a significant interaction of image type and electrode along the midline between 108 and 120 ms, $F = 23.31$, $p<.0001$, $\epsilon = 0.6715$. These occipital differences reflected processes that are likely similar to the parallel N1(00) difference between words and objects.

P150 Effects

Overall. Mean amplitude analyses between 125 and 175 ms indicated that the "P150" (Figure 1B), a broadly distributed positivity maximal around Cz, was modulated by image type, as indicated in the overall ANOVA comparing all six image types, lateral $F(5,55) = 30.62$, $p < .0001$; midline $F(5,55) = 25.69$, $p < .0001$. The P150 was not modulated by task, as the main effects of task, lateral $F < 1.5$; midline $F < 1.6$, and the interactions of task with image type were not significant, lateral $F <$ 1.2; midline $F < 1$; in subsequent analyses, therefore, the data were collapsed across task.

Word versus objects. Regarding processing differences between linguistic and nonlinguistic images, the amplitude of the P150 was much larger to words than to objects, lateral $F(1,11) = 81.10$, $p < .0001$; midline $F(1,11) = 76.52$, $p < .0001$.

Other image type comparisons. The amplitude of this positivity to icon strings was intermediate to that between words and objects; the main effects of image type were significant for both comparisons between words and icons, lateral $F(1,11) = 12.22$, $p < .01$; midline $F(1,11) = 23.33$, $p < .001$, and between objects and icons, lateral $F(1,11) = 17.64$, $p < .01$; midline $F(1,11) = 13.36$, $p < .01$. Among wordlike images, word, nonword, and pseudofont string P150s were indistinguishable (lateral $F < 1$; midline $F < 2$); even the interaction of image type with electrode was not reliable (lateral $F < 1$; midline $F < 3$). Although the comparison between objects and pseudoobjects showed no significant main effects of image type (lateral $F < 2$; midline $F < 1$), there were significant interactions of image type with electrode, lateral $F(10,110) = 11.27$, $p < .001$, $\epsilon = 0.2407$; midline $F(3,33) < 3.3$, suggesting that they may differ at some lateral sites.

Aspect Ratio and String Length

Regarding aspect ratio, there were no significant effects in any time window, including during the P150 between 125 and 175 ms (lateral $F < 1$; midline $F < 1$). Similarly, regarding string length, at lateral electrodes there were also no significant effects in any time window; this lack of effect was the case even during the P150 between 125 and 175 ms when there was no main effect of string length ($F < 1$) and no interaction with electrode ($F < 2$). However, along the midline and particularly at Oz between 100 and 175 ms, the ERPs were somewhat modulated by string length. Between 100 and 125 ms, this effect was indicated by an almost significant interaction between string length and electrode along the midline, $F(6,66) = 2.63$, $p = 0.0699$, $\epsilon = 0.4228$, and a main effect of string length when the Oz site was analyzed alone, $F(2,22) = 8.15$, $p = 0.0028$, $\epsilon = 0.7585$. Between 125 and 175 ms, there was an interaction of string length with electrode along the midline, $F(6,66) = 2.98$, $p < .05$, $\epsilon = 0.4998$, and a significant main effect at Oz, $F(2,22) = 8.92$, $p < .002$, $\epsilon = 0.8221$,[2] but not at other midline sites, 23 $F < 1.1$, Cz $F < 1$, Fpz $F < 2.5$. It is important to note that it was only where minimal or no P150 activity was visible at the midline occipital site, Oz, that the ERPs showed some slight sensitivity to string length.

Experiment 2

Onset Latency

Words versus objects. We found that word and object ERPs diverged first occipitally but in this case around 90 ms, slightly earlier than in Experiment 1, which involved fewer participants. In particular, there were significant interactions of image type at lateral electrodes, $df(10,180)$, between 84 and 96 ms (lateral $F = 6.19$, $p < .005$, $\epsilon = 0.1521$; midline $F < 5$) and 96 and 108 ms (lateral $F = 6.38$, $p = .0054$, $\epsilon = 0.1581$; midline $F < 5$). For electrode pair contrasts, main effects of image type at only lateral occipital sites (OL,OR) were significant between 84 and 96 ms

---

[1] Significant if $p \leq .0250$, $df(1,11)$.

[2] Significant if $p \leq .0125$, $df(1,11)$.

($F = 13.92$, $p < .002$) and almost significant between 96 and 108 ms ($F = 8.47$, $p = .0093$).[3] Between 108 and 120 ms, unlike the null effect in Experiment 1, a lateral difference was also indicated by a significant interaction of image type with electrode (lateral $F = 6.88$, $p = .0035$, $\epsilon = 0.1599$; midline $F < 5$).

More important, around 125 ms, the ERPs began to differentiate strongly between words and objects. This difference was indicated by significant main effects of image type between 120 and 132 ms, lateral $F(1,18) = 78.18$, $p < .0001$; midline $F(1,18) = 44.81$, $p < .0001$, that interacted with electrode, lateral $F(10,180) = 11.12$, $p < .001$, $\epsilon = 0.1677$; midline $F(3,54) = 10.05$, $p < .001$, $\epsilon = 0.5264$. In contrast to earlier differences, this effect occurred everywhere except occipitally; there were no significant main effects on occipital electrode contrasts (OL,OR $F < 3$; 23,Oz $F < 5$). This finding suggests that by 125 ms the P150 differences between words and objects have begun, whereas the lateral occipital N150 differences have not.

Words versus faces. This comparison is another between linguistic and nonlinguistic images. Similarly, the ERPs to words and faces indicated that they first diverge around 90 ms occipitally, but perhaps also at a few other lateral sites. This divergence was indicated by significant interactions of image type and electrode between 84 and 96 ms, lateral $F(10,180) = 12.99$, $p < .0001$, $\epsilon = 0.2739$; midline $F(3,54) = 9.62$, $p < .001$, $\epsilon = 0.6846$, and 96 and 108 ms, lateral $F(10,180) = 11.43$, $p < .0001$, $\epsilon = 0.2674$; midline $F < 3.2$; main effects of image type were not significant between 84 and 96 ms (lateral $F < 3$, midline $F < 1$) but were significant by 96 and 108 ms, lateral $F(1,18) = 6.43$, $p < .05$; midline $F < 3.1$. Electrode pair contrasts pinpointed these effects. There were main effects at electrode pair OL,OR between 84 and 96 ms ($F = 15.82$, $p = .0009$);[4] although main effects only approached significance at OL,OR between 96 and 108 ms ($F = 6.59$, $p = .0194$) (see Note 4), they were significant at a few anterior lateral sites, both between 84 and 96 ms (2,10, 3,9) (see Note 4) and 96–108 ms (11,20; 3,9; 12,19; 21,25).[5] More important, there were no significant differences between words and faces when the P150 began to develop; the main effects of image class were not reliable between 108 and 120 ms (lateral $F < 1$; midline $F < 1$) or between 120 and 132 ms (lateral $F < 1$; midline $F < 2$), nor were interactions of image class with electrode reliable between 108 and 120 ms (lateral $F < 3.7$; midline $F < 1$) or between 120 and 132 ms (lateral $F < 3$; midline $F < 2$). Thus, before 108 ms, early occipital differences between words and faces mirror those between words and objects; this finding is consistent with differences in visual properties between words versus faces and other objects. After 108 ms, these differences disappear for a while during the P150.

Faces versus objects. In contrast, the comparisons between ERPs to faces and objects, which are visually similar nonlinguistic images, revealed no significant differences before 120 ms, that is, until P150 effects begin. Specifically, there were no reliable main effects of image type between 108 and 120 ms (lateral $F < 1$; midline $F < 1$) or during earlier time windows (all lateral $F < 6.2$; all midline $F < 5$). More important, there were no significant interactions of image type with electrode between 108 and 120 ms (lateral $F < 1$; midline $F < 3$) and during earlier time windows

(all lateral $F < 3$; all midline $F < 2.4$). After 120 ms, the ERPs began to differentiate strongly between faces and objects, as indicated by significant main effects of image type between 120 and 132 ms, lateral $F(1,18) = 12.11$, $p < .003$; midline $F(1,18) = 11.69$, $p < .004$. However, this difference was not reliable occipitally, where the N150 appears; electrode contrasts showed no significant main effects of image type at occipital sites (OL,OR $F < 1.1$; Oz $F = 1$), which stands in contrast to the significant effects at more anterior sites (Fpz,Cz $F = 18.08$, $p < .001$; all other sites $F > 6.9$);[6] there were also no interactions of image type with hemisphere at occipital sites (OL,OR $F < 6.4$). Thus, ERPs to faces compared with ERPs to other objects do not diverge until 125 ms, but at this time the lateral occipital N150 effects are not reliable, as was also observed between words and objects.

Other image type comparisons. Unlike Experiment 1, there were no significant differences between objects and pseudoobjects until after at least 175 ms (see also P150 analyses). However, the pattern of other image type effects did replicate. Specifically, ERPs to wordlike images did not differ until after 175 ms, and the trend in Experiment 1 for ERPs to wordlike images to diverge from icon strings, a midline Oz N1(00) effect, was significant but started earlier; a midline main effect of image type between words and icon strings was significant at 96–108 ms, $F(1,18) = 10.96$, $p = .0039$, and midline interactions of image type and electrode were significant at 108–120 ms, $F(3,54) = 6.78$, $p = .0030$, $\epsilon = 0.7180$, and almost significant at 120–132 ms, $F = 3.49$, $p = .0412$, $\epsilon = 0.6757$. Within this period, occipital ERPs to icon strings and objects diverged, as indicated by significant interactions of image type and electrode between 108 and 120 ms, lateral $F(10,180) = 8.39$, $p = .0012$, $\epsilon = 0.1758$; midline $F(3,54) = 13.75$, $p = .0012$, $\epsilon = 0.4213$. This difference likely reflected the same processes as the parallel difference between words and objects, which probably reflected differences in low-level visual properties.

P150 Effects
Word versus objects. As in Experiment 1, the mean amplitude of the P150 (i.e., the positivity between 125 and 175 ms) was greater to words than to objects (Figure 3B). This finding was indicated by main effects of image type, lateral $F(1,18) = 81.81$, $p < .0001$; midline $F(1,18) = 66.81$, $p < .0001$.

Faces versus objects. The mean amplitude of the positivity between 125 and 175 ms was also larger to faces than to objects (Figure 3B). This finding was indicated by main effects of image type, lateral $F(1,18) = 42.88$, $p < .0001$; midline $F(1,18) = 56.40$, $p < .0001$. Thus, we replicated the main differences reported by others between faces and objects.

Words versus faces. The comparison between words and faces, however, was the main focus of this experiment. Thus, it is most important that the P150 amplitudes to words and faces were indistinguishable everywhere, except occipitally (Figure 3A,B). There were no main effects of image type on either mean (lateral $F < 16$; midline $F < 2$) or peak (lateral $F < 1$; midline $F < 3.7$) amplitude. However, an occipital effect was indicated by (a) significant two-way interactions of Image Type × Hemisphere for both mean amplitude, lateral $F(1,18) = 11.21$, $p < .004$, and peak amplitude,

---

[3] Significant if $p \leq .0038$, $df(1,18)$.
[4] Significant if $p \leq .0045$, $df(1,18)$.
[5] Significant if $p \leq .00625$, $df(1,18)$.

[6] Significant if $p \leq .0170$, $df(1,18)$.

Figure 3. (**A**) Experiment 2: Normalized P150 peak (152–156 ms) voltage maps are very similar between words and faces (for map calculation, see Ganis, Kutas, & Sereno, 1995). (**B**) Grand-average ERPs from −150 ms to 350 ms recorded in Experiment 2. P150 amplitude to words and faces are indistinguishable, except occipitally, where an N150 instead is apparent. Shading indicates the P150 difference between faces and objects, which resembles that between words and objects.

lateral $F(1,18) = 11.52$, $p < .004$; (b) marginally significant three-way interactions of Image Type × Hemisphere × Electrode for both mean amplitude, lateral $F(10,180) = 4.10$, $p < .05$, $\epsilon = 0.1858$, and peak amplitude, lateral $F(10,180) = 4.80$, $p < .05$, $\epsilon = 0.1908$; and (c) marginally significant two-way interactions of Image Type × Electrode for both mean amplitude, lateral $F(10,180) = 5.67$, $p < .05$, $\epsilon = 0.1356$; midline $F(3,54) = 5.93$, $p < .05$, $\epsilon = 0.4200$, and peak amplitude, lateral $F(10,180) = 5.42$, $p < .05$, $\epsilon = 0.1395$; midline $F(3,54) = 6.17$, $p < .05$, $\epsilon = 0.4271$. Specifically, words elicited greater lateral occipital negativity than faces, an N150.

Contrasts on electrode pairs localized the N150 to occipital sites. Specifically, with mean amplitude analyses, there were significant differences between words and faces. This finding was indicated by main effects of image type for the N150 at OL,OR ($F = 12.31$) but not for the P150 elsewhere (i.e., at 15,16 $F < 3.1$; other lateral sites, $F < 1.3$; or any midline sites, 1,Cz $F < 1$; 23,Oz $F < 6.5$). Similarly, with peak amplitude analyses, main effects of image type were significant for the N150 at OL,OR ($F = 14.07$) but not significant for the P150 elsewhere (at 15,16 $F < 4.6$; other lateral $F < 1.5$; or midline sites 1,Cz $F < 1$; 23,Oz $F = 8.58$) (Note 3). Although the N150 effect at lateral occipital sites (OL,OR) appears to be larger over the left hemisphere, neither the main effect of hemisphere (both mean and peak amplitude: $F < 1$) nor the Image Type × Hemisphere interaction was significant (mean amplitude $F < 7.5$; peak amplitude $F < 8$) (see Note 3).

Peak latency: P150 versus N150. Because it has the same peak latency, the N150 may indicate a polarity inversion of the P150. For words, faces, and objects, there were no significant main effects of peak latency for comparisons between the lateral occipital N150 (at OL,OR) and the P150 measured at nearby sites, either temporal, 4,8 (words $F < 1$; faces $F < 1.2$; objects $F < 1$) or parietal, 14,17 (words $F < 1$; faces $F < 3$; objects $F < 1$). However, the lateral occipital N150 also reflects processes that are somewhat distinct from those underlying the P150. This suggestion arose from significant interactions of peak latency and hemisphere only for words with both the temporal comparison, that is, OL,OR N150 versus 4,8 P150 (words $F = 19.63$, $p = .0003$; faces $F < 3$; objects $F < 4.4$), and the parietal comparison, that is, OL,OR N150 versus 14,17 P150 (words $F = 6.09$, $p = .0239$; faces $F < 4$; objects $F < 1$), and the N150 difference between words and faces previously described that was not found either at the P150 peak or between 125 and 175 ms.

Other image type comparisons. As in Experiment 1, mean amplitude analyses between 125 and 175 ms indicated that the positivity to icon strings was intermediate to that for words and objects; this finding was indicated by significant main effects of image type for both words versus icons, lateral $F(1,18) = 7.6$, $p < .05$; midline $F(1,18) = 9.29$, $p < .01$, and objects versus icons, lateral $F(1,18) = 49.09$, $p < .0001$; midline $F(1,18) = 31.25$, $p < .0001$. The positivities to words, nonwords, and pseudofont

strings were indistinguishable, as indicated by no significant main effects of wordlike image type (lateral $F < 1$; midline $F < 2$). Although the midline ANOVA suggested a significant interaction of wordlike image with electrode, lateral $F < 2.2$; midline $F(6, 108) = 5.27$, $p < 0.005$, $\epsilon = 0.4748$, this interaction was not observed in Experiment 1. Object versus pseudoobject differences did not replicate because in this experiment there were no significant main effects of objectlike image type (lateral $F < 1$; midline $F < 2.1$) or interactions with electrode (lateral $F < 2.8$; midline $F < 1$).

Difference Wave Analyses

Onset latency. The analyses on raw ERPs indicated that, although the P150 peaks between 152 and 156 ms, the onset of ERP differences between words and objects and between faces and objects occurs between 120 and 132 ms, or around 125 ms. We further evaluated this onset latency result by comparing the shape of the scalp distributions of normalized difference ERPs. All ANOVAs contrasted these measures at the P150 peak (152–156 ms) to those for other 4-ms time windows before the peak, starting from 120 ms and ending at 152 ms. For words minus faces during this entire time, there were no significant differences in the shape of the scalp distributions, all $Fs(25, 450) < 1.6$, $ps > .20$. For faces minus objects, the differences in the shape of the scalp distributions approached significance from 120 to 128 ms ($\alpha = 0.0125$; 120–124 ms, $F = 3.88$, $p = .0135$, $\epsilon = 0.1295$; and 124–128 ms, $F = 2.9$, $p = .0396$, $\epsilon = 0.1227$), and for words minus objects, differences in the shape of the scalp distributions also approached significance from 120 to 124 ms ($F = 2.60$, $p = .0601$, $\epsilon = 0.1332$). However, the more important finding was that after 128 ms for both words minus objects and for faces minus objects, the shape of the scalp distributions were indistinguishable from those at the P150 peak (all $Fs < 2.1$, $ps > .10$). This observation supports our working hypothesis that the effects observed at the P150 peak begin as early as approximately 125 ms.

Temporally overlapping later ERPs. There were significant interactions between time window and electrode for normalized difference ERPs (words minus objects, faces minus objects, or words minus faces) at the P150 peak versus later time epochs (Table 1). These interactions were observed with both kinds of ANOVAs that included either all electrodes or all but the occipital sites. Thus, the shape of the scalp distribution of the P150 peak differs from later ERP components that are active within 650 ms, such as N200, P300, N400, and P600 (Figure 4).

DISCUSSION

Overview

Across two ERP experiments, we compared the early responses of the human brain to a variety of linguistic visual images, such as words, and nonlinguistic objects, such as animals, tools, and faces. We focused on the nature and onset of the first perceptual processing differences between linguistic and nonlinguistic visual images that reflect more than differences in simple, low-level image properties.

N1(00) and P1(00)

The first ERP differences between linguistic and nonlinguistic visual images were observed focally at occipital sites on the N1(00)

Table 1. *F Values for Interactions Between Time Window and Electrode on Measures of the Shape of the Scalp Distribution of Normalized Difference ERPs for the P150 Peak (152–156 ms) Versus Later Time Windows*

	152–156 ms vs.				
	175–250	250–350	350–450	450–550	550–650
Words vs. objects					
All sites					
F	9.92**	7.81**	8.19*	12.69**	6.48*
ϵ	0.1405	0.1490	0.1146	0.1226	0.1349
All but occipital					
F	10.09**	6.80*	7.37*	14.63**	5.44*
ϵ	0.1424	0.1540	0.1299	0.1407	0.1624
Faces vs. objects					
All sites					
F	18.78**	21.27**	21.94**	13.79**	10.83**
ϵ	0.1895	0.1690	0.1666	0.1620	0.1960
All but occipital					
F	17.27**	15.64**	14.76**	7.63*	7.49**
ϵ	0.1893	0.1755	0.1936	0.1578	0.1998
Words vs. faces					
All sites					
F	12.06**	16.30**	17.83**	13.95**	10.66**
ϵ	0.1154	0.1183	0.1186	0.1205	0.1293
All but occipital					
F	14.68**	20.85**	17.39**	11.42**	12.50*
ϵ	0.1868	0.1817	0.1900	0.1864	0.1480

Note: Significant interactions between time window and electrode revealed that the shape of the scalp distribution of image type effects differ between the P150 peak and all later time epochs (see Figure 4). All sites = 26 lateral and midline electrodes analyzed; all but occipital = 23 lateral and midline electrodes analyzed without the three occipital sites (OL, OR, Oz). All sites: $df(25, 450)$; all but occipital: $df(22, 396)$.
*$p < .001$, **$p < .0001$.

at Oz or the P1(00) at OL, OR between 90 and 110 ms. This time is shortly after activation of striate cortex between 50 and 80 ms (Clark, Fan, & Hillyard, 1995). In particular, by 90 ms, the occipital ERPs differentiated between words and other wordlike strings versus faces (Experiment 2) and other objects (both experiments). This divergence is consistent with the inherent visual dissimilarity between words and faces and between words and objects. In contrast, as predicted by their visual similarity, early occipital differences were not found between faces and other objects or among wordlike strings (i.e., words, nonwords, and pseudofont strings). The occipital differences probably do not reflect neural operations that have become specialized for aspects of linguistic processing per se. Rather, they seem to reflect differences between words versus faces and other objects in low-level visual properties, such as spatial frequency spectra (De Valois & De Valois, 1988) or perhaps spatial dimensions [e.g., string length seemed to modulate the N1(00) at Oz]. These first N1(00) and P1(00) effects were very small compared with later effects.

P150

The next image type effects were indexed at least initially by a broadly distributed, bilateral positivity that was maximal at the vertex site, Cz, around 155 ms. Identification of these effects with
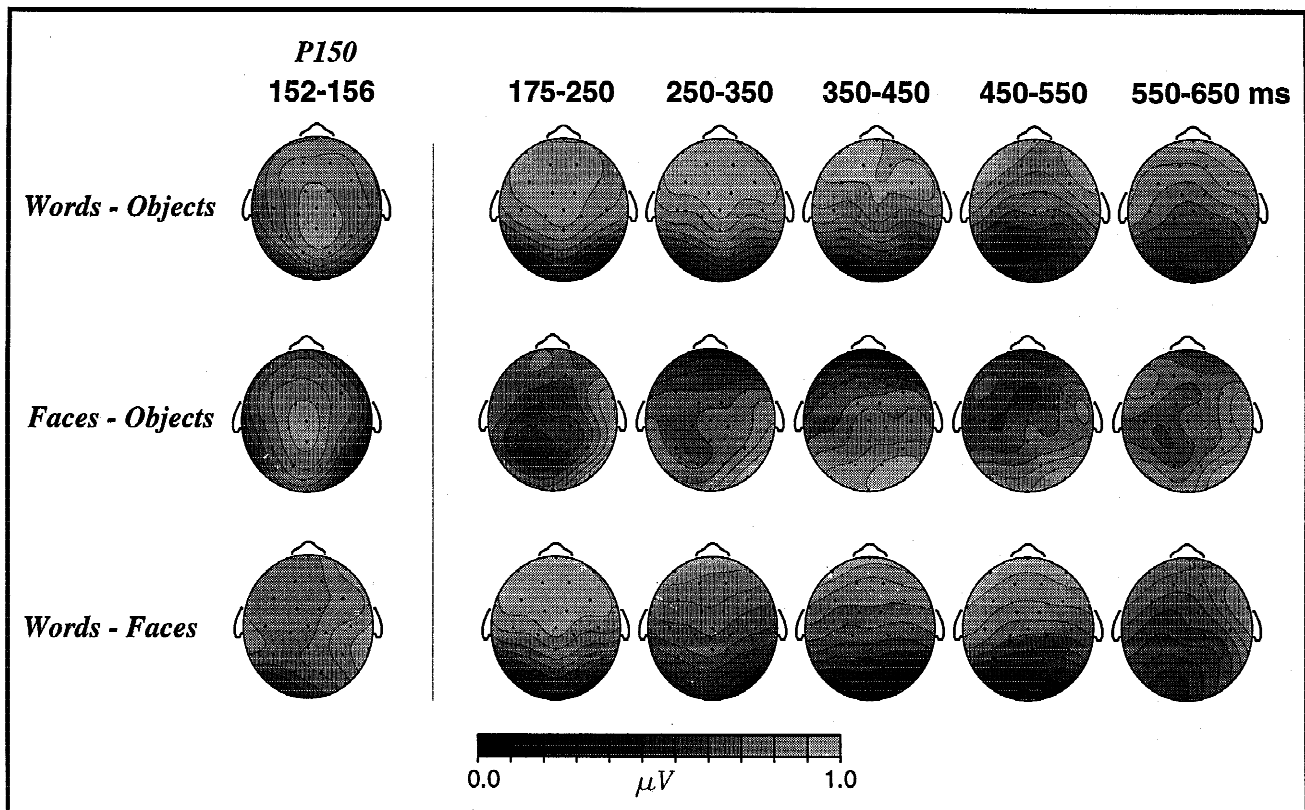
Figure 4. Normalized voltage maps showing the scalp distribution of difference waves for words minus objects, faces minus objects, and words minus faces, respectively from the top. The shapes of the scalp distributions of the normalized difference waves at the P150 peak (152–156 ms) differ markedly from those measured in later time windows that correspond to when N200 (175–250 ms), P300 (250–350 ms), N400 (350–450 ms), and P600 (450–550 and 550–650 ms) slow waves are observed (see Table 1 for statistics).

the main ERP peak, the "P150," where they appeared to be observed initially, was further supported by detailed comparisons with both earlier and later ERPs.

Image Type Selectivity

The P150 was much more positive to wordlike than to objectlike images and intermediate to icon strings, which are both wordlike and objectlike. The P150s to wordlike images, namely words, nonwords, and pseudofont strings, were indistinguishable; this coarse stimulus selectivity indicates that the underlying neural operations have relatively limited representation capabilities, as expected at early stages in the visual processing hierarchy (Kobatake & Tanaka, 1994). Because the P150 did not vary with string length or aspect ratio, low-level visual properties cannot explain the pattern of its modulation. Overall, these findings are consistent with those from a study of native Japanese speakers in which a P150 to Japanese words (single kanjis) was larger than that to line drawings of objects (Shimoyama et al., 1992); kanji words are ideographs, resembling line drawings of objects in visual and perceptual properties. Thus, in general, P150 activity seems to be engaged preferentially by images resembling any well-learned, linguistic visual pattern.

Alternatively, the P150 may not be language specific but rather elicited by images resembling any well-learned perceptual category of visual patterns. Indeed, there is a positivity peaking at around 150–200 ms that is largest to any face depiction, known as

the "P150" (Bötzel & Grüsser, 1989), the "P200" (Potter & Parker, 1989), or the "vertex positive peak" (Jeffreys, 1989). Fragmented, schematic, mildly distorted, and illusory faces still elicit larger P150s than do other categories of objects (Bötzel & Grüsser, 1989; George, Evans, Fiori, Davidoff, & Renault, 1996; Jeffreys, 1989, 1993; Jeffreys & Tukmachi, 1992; Jeffreys, Tukmachi, & Rockley, 1992). Moreover, any factor that impairs classification of an image as a face likewise reduces P150s, such as degradation, orientation inversion, or fixation off the eyes (Jeffreys, 1989, 1993; Jeffreys et al., 1992).

To explore the relationship between word and face P150s, in Experiment 2 we compared them directly in the same individuals using visually matched images. We found that the amplitude and distribution of the P150s to words and faces were remarkably similar, while also replicating the smaller P150 to other categories of visual objects. Our effects were not a consequence of processing a single category (faces or words) versus several categories of other objects; the same pattern of differences between faces and other objects have been observed in studies in which faces were compared with only two categories of other objects, such as flowers and leaves (Bötzel, Schulze, & Stodieck, 1995).

Functional Significance

Thus, the P150 is neither merely a "face-responsive" ("face-specific" or "face-related") potential (Bötzel & Grüsser, 1989; Jeffreys, 1989; Seeck & Grüsser, 1992) nor merely a "word-

selective or language-specific potential. Therefore, we propose that P150 amplitude is sensitive to the statistical regularities within well-learned, visual image categories. Visually, faces resemble objects more than they resemble words. However, faces resemble words more in terms of the amount of experience people have discriminating among, or identifying, particular instances of them; this task, which has strong behavioral relevance, is performed more frequently on both words and faces than on other visual objects. Thus, modulation of this early positivity may reflect the cumulative experience people have discriminating between exemplars within particular categories of visual images. Activated regions of primary motor or visual cortex expand after skill learning (Karni et al., 1995; Ungerleider, 1995). Perhaps learning operations recruit synchronously active neurons into visual representations, resulting in increased P150 amplitude. Extensive training on novel, artificial, visual stimuli can produce behavioral effects resembling those obtained with faces (Gauthier & Tarr, 1997). Our hypothesis predicts that P150 amplitude to such artificial stimuli should increase with learning.

The ethological utility of early perceptual categorization has been expressed by Ullman (1996): "If the image can be classified, for example, as representing a face, without identifying the individual face, then subsequent processing stages can be directed to specific face models, rather than to models of other 3-D objects" (pp. 163–164). Jeffreys (1996) proposed that the processes indexed by the face P150 are designed for very rapid detection of suddenly fixated faces. Based on our results, we speculate that the perceptual operations indexed by the P150 may function to rapidly orient attention to images with a well-learned salience, such as faces and words, thereby facilitating subsequent processing of these ethologically important visual stimuli (e.g., Schendan, Kanwisher, & Kutas, 1997).

Onset Latency

The onset of the P150 may thus index the earliest point in processing when the human brain categorizes images based on learned, higher order, perceptual features, rather than merely simple visual features, such as spatial frequency. The onset latency was determined by comparisons between image types on raw ERPs, and, most important, by comparisons between the P150 peak and earlier time windows of the shape of the scalp distribution of normalized difference waves (e.g., ERPs to words minus those to objects across the head that are normalized to adjust for amplitude differences; McCarthy & Wood, 1985). The former analyses on raw ERPs suggested that the P150 differences between words and faces versus other objects begin between 120 and 132 ms. The latter analyses revealed that, between 125 and 152 ms, the shape of the scalp distribution is the same as that at the P150 peak. This finding suggests that the ERP differences between words and faces versus other objects that occur between 125 ms and the P150 peak reflect modulations of the same neural generators. We are therefore confident that the onset of differences around 125 ms indicates the beginning of the perceptual discrimination indexed by word and face P150s relative to other objects.

Relationship With Later ERPs

Modulation of P150 amplitude cannot be attributed solely to other temporally overlapping ERP components. For example, broad scalp N200 enhancement has been reported for emotionally negative novel faces relative to repeated faces and both novel and repeated words (Deldin, 1996; Miller, 1996). The shape of the scalp distribution is a marker of the configuration of a set of underlying neural

generators. If the P150 effects are due entirely to modulation of an incoming scalp N200, then there should be no difference in the shape of the scalp distribution between the P150 and the N200 difference waves. This line of reasoning applies also to other late slow wave components such as P300, N400, and P600.

On the contrary, as early as 175–250 ms, the shape of the scalp distribution of the difference waves (namely for the scalp N200) were markedly different from those at the P150 peak (i.e., 152–156 ms). Furthermore, the shape of the scalp distributions for later difference waves (namely P300, N400, and P600) also differed markedly from those at the P150 peak. Thus, the P150 indexes neural operations that are either somehow distinct or activated differentially from those underlying later ERP components. Consequently, temporally overlapping N200, P300, N400, or P600 components cannot be the sole explanation for P150 effects. Of course, some contribution from these other ERPs to effects within the P150 latency range cannot be ruled out entirely. However, this report focused on the P150 as an index of the initial time when the ERPs diverge in response to learned higher order information in visual images, namely perceptual categorization; we focused on the P150 as the component that captures at least the onset of this divergence.

Neural Generators

The scalp distributions of word and face P150s were similar, suggesting that the neural generators are the same or nearby. The scalp distributions were broad, which is consistent with neural activity from deep within the brain. However, no consensus has been reached on the location of a face P150 generator, which has been localized to widely differing brain areas, including posterior fusiform gyrus (Allison, McCarthy, et al., 1994; Bötzel & Grüsser, 1989; Jeffreys, 1989; Sams, Hietanen, Hari, Ilmoniemi, & Lounasmaa, 1997), superior temporal sulcus (Jeffreys, 1989), inferior occipitotemporal junction (Lu et al., 1991), and amygdala, hippocampus, or cingulate gyrus (Bötzel & Grüsser, 1989; Bötzel et al., 1995). This lack of consensus illustrates the shortcomings of current methods of source localization from scalp distribution alone, especially for deep generators (Srebro, Oguz, Hughlett, & Purdy, 1993). The functional characterization of ERPs is therefore an essential constraint on source localization. Scalp P150s to words and faces seem to reflect functionally similar processes that accomplish the rapid detection of categories of frequent and salient visual stimuli. We hypothesize that such processes may be embodied in distinct but spatially adjacent neural populations.

In particular, similar to scalp P150s, intracranial potentials, namely N200s, (a) are larger to words and faces than to other objects, (b) do not reliably differentiate words from nonwords, and (c) exhibit similar face inversion effects (Allison, Ginter, et al., 1994; Allison, McCarthy, et al., 1994; Jeffreys, 1989, 1993, 1996; Nobre, Allison, & McCarthy, 1994). Accordingly, we propose that the scalp P150 largely reflects activation of posterior fusiform gyrus where intracranial N200s (150–200 ms)[7] seem to reflect visual perceptual categorization, namely prelexical analysis of let-

---

[7] Apparent discrepancies between the peak latencies of the scalp P150 and intracranial N200 may be accounted for by any or all of several factors. (a) Although Allison and collaborators referred to the intracranial component as an "N200," the latency of this component ranges from 150 to 200 ms. (b) The stimuli in intracranial studies are gray-scale photographic images that may elicit somewhat later N200s than the simpler line drawings we used. (c) Although the average age of our participants was younger than 25 years, that for participants in the intracranial studies was older than 40; increases in the latency of ERPs have been observed with aging.

terstrings or face detection (Allison, McCarthy, et al., 1994). For faces, some authors have suggested that the scalp P150 may reflect activation of the posterior fusiform gyrus N200 (Allison, McCarthy, et al., 1994; Bötzel & Grüsser, 1989; Jeffreys, 1989, 1996; Sams et al., 1997). Our experiments are the first to implicate a similar posterior fusiform gyrus generator for the scalp P150 to words (and wordlike images).

The location of the N200 is consistent with this proposal. The intracranial N200 is a negative field (with mastoids as references) across the base of the brain (Allison, McCarthy, et al., 1994); the N200 would produce a diffuse positivity over the upper half of the head (Wood & Wolpaw, 1982). Neural generators of letterstring and face N200s are adjacent but segregated within the posterior fusiform gyrus (Allison, McCarthy, et al., 1994), which would result in scalp potentials with highly similar amplitudes and distributions. Although letter string N200s have been observed only in posterior fusiform gyrus, face N200s have occasionally also been observed several centimeters away in the inferotemporal cortex (Allison, Ginter, et al., 1994; Allison, McCarthy, et al., 1994; Nobre et al., 1994); an inferotemporal generator may produce lateral posterior differences between faces and words.

This picture of letter string and face processing is supported by functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) research. Posterior fusiform gyrus is activated by both letter strings and faces (Haxby et al., 1994; Kanwisher, Chun, McDermott, & Ledden, 1996; Price, Wise, & Frackowiak, 1996; Puce, Allison, Asgari, Gore, & McCarthy, 1996; Sergent, MacDonald, & Zuck, 1994). However, the area of activation is smaller for letter strings, being concentrated instead in occipitotemporal and inferior occipital sulci (Puce et al., 1996), which are also activated during face perception (Haxby et al., 1994). Puce et al. (1996) suggested that the temporal integration of activity in fMRI and PET studies may have obscured the posterior fusiform gyrus activation that is specific to letterstrings.

Activation of posterior fusiform gyrus that overlaps temporally with that of occipitotemporal and inferior occipital sulci and that of inferotemporal gyrus generators is consistent with the complex pattern of scalp ERP effects observed in our experiments, especially posteriorly. At occipital sites, the P150 was larger to faces than to words, which is consistent with the greater extent of posterior fusiform gyrus activation in fMRI to faces relative to letter strings (Puce et al., 1996). The P150 appears to reverse in polarity at lateral occipital sites, where an N150 is apparent. The N150 was largest to words, which is consistent with reports of a letter-string "N2" (Nobre & McCarthy, 1994), perhaps reflecting occipitotemporal and inferior occipital sulci activation adding to posterior fusiform gyrus activation. We believe that the N150 reflects processes that are similar to but somewhat distinct from those of the P150 because the P150 effects begin earlier and the pattern of

effects on the two differ. Specifically, at the P150 onset between 120 and 132 ms, there are differences in P150 amplitude between faces and other objects and between words and other objects that are absent with the occipital N150. Moreover, mean and peak amplitudes between 125 and 175 ms to words and faces differ at the N150 but not at the P150. Thus, P150 and N150 reflect somewhat distinct processes. However, because some processes may be shared, our discussion of P150 effects may apply also to the N150, albeit to a lesser extent.

Although hemispheric asymmetry of the N150 was apparent but not reliable in our results, an N150 to words has been reported to be larger over the left hemisphere (Nobre & McCarthy, 1994), whereas an N150 to faces has been reported to be larger over the right hemisphere (Bentin, Allison, Puce, Perez, & McCarthy, 1996). Such hemispheric asymmetry of N150 effects is consistent with the idea that left occipitotemporal and inferior occipital sulci are more strongly engaged by letter strings, whereas right occipitotemporal and inferior occipital sulci are more strongly engaged by faces or eyes (Bentin et al., 1996; Puce et al., 1996). That hemispheric asymmetry has been observed for the N150 but not for the P150 provides further evidence that the N150 and the P150 reflect the activity of somewhat distinct neural generators.

Conclusions

By directly comparing words, faces, and other objects, these studies showed that by as early as 125 ms the human brain differentiates between images resembling well-learned categories of visual objects (e.g., words and faces) versus those resembling other, less frequently recognized objects (e.g., animals and tools). This level of visual perceptual categorization is characterized at least initially as a broadly distributed scalp positivity, the P150. The 125-ms onset of effects indexed by the P150 thus seems to demarcate one of the earliest points, if not the earliest, when perceptual categorization of visual objects begins. We suggest that P150 amplitude reflects, among other things, long-term experience with the statistical regularities of visual input. Experience may tend to have the greatest effect during more demanding and behaviorally relevant visual tasks, such as recognizing instances of particular object categories; reading and identification of particular people are two such tasks for words and faces, respectively. The stimulus specificity of the image representations thereby developed are constrained by the encoding capabilities of the cortical areas involved. We propose that the main generator of the scalp P150 to words and faces is the posterior fusiform gyrus, in which letter-string and face N200s have been recorded intracranially. Thus, the scalp P150 may be used for noninvasive studies of higher order visual processing within early extrastriate cortex.

REFERENCES

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M., & Spencer, D. D. (1994). Face recognition in human extrastriate cortex. *Journal of Neurophysiology*, *71*, 821–825.

Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex*, *4*, 544–554.

Amir, A. (1994). Uniqueness of the generators of brain evoked potential maps. *IEEE Transactions on Biomedical Engineering*, *41*, 1–11.

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

Bötzel, K., & Grüsser, O. J. (1989). Electric brain potentials evoked by pictures of faces and non-faces—A search for face-specific EEG-potentials. *Experimental Brain Research*, *77*, 349–360.

Bötzel, K., Schulze, S., & Stodieck, S. R. G. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental Brain Research*, *104*, 135–143.

Clark, V. P., Fan, S., & Hillyard, S. A. (1995). Identification of early visual evoked potential generators by retinotopic analyses. *Human Brain Mapping*, *2*, 170–187.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.

Deldin, P. J. (1996). *Memory bias in major depression: The P300 connection*. Unpublished doctoral dissertation, University of Illinois, Champaign.

Farah, M. J. (1994). Specialization within visual object recognition: Clues from prosopagnosia and alexia. In M. J. Farah & G. Ratcliff (Eds.), *The neuropsychology of high-level vision: Collected tutorial essays* (Carnegie Mellon Symposia on Cognition, pp. 133–146). Hillsdale, NJ: Erlbaum.

Ganis, G., Kutas, M., & Sereno, M. I. (1995). Freeing the ERPs: Freeware for high quality spatial map construction and presentation [abstract]. *Psychophysiology*, *32*(Suppl. 1), S33.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for common sense—An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*, 89–106.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673–1682.

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

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex—A PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience*, *14*, 6336–6353.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*, 800–802.

Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, *78*, 193–202.

Jeffreys, D. A. (1993). The influence of stimulus orientation on the vertex positive scalp potential evoked by faces. *Experimental Brain Research*, *96*, 163–172.

Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition*, *3*, 1–38.

Jeffreys, D. A., & Tukmachi, E. S. A. (1992). The vertex-positive scalp potential evoked by faces and by objects. *Experimental Brain Research*, *91*, 340–350.

Jeffreys, D. A., Tukmachi, E. S. A., & Rockley, G. (1992). Evoked potential evidence for human brain mechanisms that respond to single, fixated faces. *Experimental Brain Research*, *91*, 351–362.

Kanwisher, N., Chun, M. M., McDermott, J., & Ledden, P. J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research*, *5*, 55–67.

Karni, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature*, *377*(6545), 155–158.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*, 856–867.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Lu, S. T., Hamalainen, M. S., Hari, R., Ilmoniemi, R. J., Lounasmaa, O. V., Sams, M., & Vilkman, V. (1991). Seeing faces activates three separate areas outside the occipital visual cortex in man. *Neuroscience*, *43*(2–3), 287–290.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.

Miller, G. A. (1996). How we think about cognition, emotion, and biology in psychopathology. *Psychophysiology*, *33*, 615–628.

Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature*, *372*(6503), 260–263.

Nobre, A. C., & McCarthy, G. (1994). Language-related ERPs: Scalp distributions and modulation by word type and semantic priming. *Journal of Cognitive Neuroscience*, *6*, 233–255.

Nunez, P. L. (1981). *Electric fields of the brain: The neurophysics of EEG*. New York: Oxford University Press.

Potter, D. D., & Parker, D. M. (1989). Electrophysiological correlates of facial identity and expression processing. In J. R. Crawford & D. M. Parker (Eds.), *Clinical and experimental neuropsychology* (pp. 137–150). New York: Plenum Press.

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature*, *253*(5491), 437–438.

Price, C. J., Wise, R. J. S., & Frackowiak, R. S. J. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex*, *6*, 62–70.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures—A functional magnetic resonance imaging study. *Journal of Neuroscience*, *16*, 5205–5215.

Rohrbaugh, J. W., Donchin, E., & Eriksen, C. W. (1974). Decision making and the P300 component of the cortical evoked response. *Perception and Psychophysics*, *15*, 368–374.

Sams, M., Hietanen, J. K., Hari, R., Ilmoniemi, R. J., & Lounasmaa, O. V. (1997). Face-specific responses from the human inferior occipito-temporal cortex. *Neuroscience*, *77*, 49–55.

Schendan, H. E., Kanwisher, N. G., & Kutas, M. (1997). Early brain potentials link repetition blindness, priming and novelty detection. *NeuroReport*, *8*, 1943–1948.

Seeck, M., & Grüsser, O. J. (1992). Category-related components in visual evoked potentials: Photographs of faces, persons, flowers and tools as stimuli. *Experimental Brain Research*, *92*, 338–349.

Sergent, J., MacDonald, B., & Zuck, E. (1994). Structural and functional organization of knowledge about faces and proper names: A positron emission tomography study. In C. Umiltá & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 203–228). Cambridge, MA: MIT Press.

Shimoyama, I., Morita, Y., Uemura, K., Kojima, Y., Nakamura, T., Nishizawa, S., Yokoyama, T., Ryu, H., Murayama, M., & Uchizono, K. (1992). Verbal versus non-verbal visual evoked potentials: Kanji versus line drawings. *Brain Topography*, *5*, 35–39.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 174–215.

Srebro, R. (1996). A bootstrap method to compare the shapes of two scalp fields. *Electroencephalography and Clinical Neurophysiology*, *100*, 25–32.

Srebro, R., Oguz, R. M., Hughlett, K., & Purdy, P. D. (1993). Functional brain imaging—Dipole localization and Laplacian methods. *Vision Research*, *33*, 2413–2419.

Stuss, D. T., Picton, T. W., Cerri, A. M., Leech, E. E., & Stethem, L. L. (1992). Perceptual closure and object identification—Electrophysiological responses to incomplete pictures. *Brain and Cognition*, *19*, 253–266.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.

Ungerleider, L. G. (1995). Functional brain imaging studies of cortical mechanisms for memory. *Science*, *270*(5237), 769–775.

Wood, C. C., & Wolpaw, J. R. (1982). Scalp distribution of human auditory evoked potentials. II. Evidence for overlapping sources and involvement of auditory cortex. *Electroencephalography and Clinical Neurophysiology*, *54*, 25–38.