Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: Evidence from parametric, test-retest, single-subject analyses

Magdalena M. Bieniek
Institute of Neuroscience and Psychology, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow, UK

Cyril R. Pernet
Brain Research Imaging Center, Division Clinical Neurosciences, University of Edinburgh, Western General Hospital, Edinburgh, UK

Guillaume A. Rousselet
Institute of Neuroscience and Psychology, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow, UK

One major challenge in determining how the brain categorizes objects is to tease apart the contribution of low-level and high-level visual properties to behavioral and brain imaging data. So far, studies using stimuli with equated amplitude spectra have shown that the visual system relies mostly on localized information, such as edges and contours, carried by phase information. However, some researchers have argued that some event-related potentials (ERP) and blood-oxygen-level-dependent (BOLD) categorical differences could be driven by nonlocalized information contained in the amplitude spectrum. The goal of this study was to provide the first systematic quantification of the contribution of phase and amplitude spectra to early ERPs to faces and objects. We conducted two experiments in which we recorded electroencephalograms (EEG) from eight subjects, in two sessions each. In the first experiment, participants viewed images of faces and houses containing original or scrambled phase spectra combined with original, averaged, or swapped amplitude spectra. In the second experiment, we parametrically manipulated image phase and amplitude in 10% intervals. We performed a range of analyses including detailed single-subject general linear modeling of ERP data, test-retest reliability, and unique variance analyses. Our results suggest that early ERPs to faces and objects are due to phase information, with almost no contribution from the amplitude spectrum. Importantly, our results should not be used to justify uncontrolled stimuli; to the contrary, our results emphasize the need for stimulus control (including the amplitude spectrum), parametric designs, and systematic data analyses, of which we have seen far too little in ERP vision research.

Keywords: phase spectrum, amplitude spectrum, object categorization, face categorization, event-related potentials, general linear model


Introduction

One of the most enduring problems in studying object categorization is to tease apart the contribution of low-level and high-level visual properties to behavioral and brain imaging results (Rousselet & Pernet, 2011; Schyns, Gosselin, & Smith, 2009; VanRullen, 2011). In particular, the contribution of image Fourier amplitude and phase spectra to behavioral performance and brain activity is one of the most studied dissociations. Whereas the amplitude spectrum contributes to the overall image appearance, the phase spectrum carries information about local image structures, such as edges and contours, because edges require phase alignment across spatial frequency components (Hansen, Farivar, Thompson, & Hess, 2008; Kovesi, 1999; Morrone & Burr, 1988). The importance of phase for object recognition has been demonstrated in studies conducted by Piotrowski and Campbell (1982) and Oppenheim and Lim (1981) who showed that, when mixing the Fourier amplitude of one image with the Fourier phase of another image, the outcome resembles its phase contributor much more than its amplitude contributor. Since then, studies using stimuli equated in amplitude spectra have
demonstrated that early object visual processing relies mostly on phase information (e.g., Allison, Puce, Spencer, & McCarthy, 1999; Jacques & Rossion, 2006; Loschky & Larson, 2008; Rousselet, Husk, Bennett, & Sekuler, 2005; Rousselet, Mace, Thorpe, & Fabre-Thorpe, 2007; Rousselet, Pernet, Bennett & Sekuler, 2008; Wichmann, Braun, & Gegenfurtner, 2006; Wichmann, Drewes, Rosas, & Gegenfurtner, 2010), with effects on brain activity starting at about 100–150 ms after the stimulus onset (Rousselet, 2012; Rousselet, Gaspar,Wieczorek, & Pernet, 2011; Rousselet, Husk, Bennett, & Sekuler, 2008; VanRullen & Thorpe, 2001).

Although these findings emphasize the importance of phase information for object categorization, they do not rule out a possible contribution of the amplitude spectrum to this process. Indeed, when images have substantially different Fourier amplitudes, the role of phase may no longer be essential (Juvells, Vallmitjana, Carnicer, & Campos, 1991). This idea is supported by the existence of computational algorithms that can efficiently classify images of natural scenes using nonlocalized or coarsely localized amplitude spectrum information (Oliva & Torralba, 2001). Furthermore, human observers can detect degradation in amplitude spectra in meaningless synthetic textures (Clarke, Green, & Chantler, 2012) or discriminate between wavelet textures using higher order statistics (Kingdom, Hayes, & Field, 2001). Hence, provided that amplitude spectrum information is available for the task at hand, human observers might be able to use it when categorizing objects and natural scenes. In particular, some studies suggest that when a stimulus is presented rapidly, the amplitude spectrum may provide a type of abstract information not obviously related to the semantic content of an image, but sufficient for its broad categorization (Crouzet & Thorpe, 2010; Honey, Kirchner, & VanRullen, 2008; Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Oliva & Torralba, 2006; VanRullen, 2006).

Alternatively to the two previous accounts, it is also plausible that object and scene categorization do not depend on phase or on amplitude alone, but on an interaction between them. For instance, categorization accuracy decreases when the amplitude of each stimulus is replaced by the average amplitude across stimuli, while retaining the original phase (Drewes, Wichmann, & Gegenfurtner, 2006). Accuracy is also affected when the amplitude is swapped within image category in an animal detection task—e.g., the amplitude spectrum of a fish is mixed with the phase spectrum of a tiger (Gaspar & Rousselet, 2009). Because swapping amplitude spectra within category should preserve their diagnostic properties in an animal detection task, this result suggests the existence of a specific relationship between phase and amplitude spectra, which when disturbed, hampers image classification.

To sum up, the importance of the amplitude spectrum in object categorization is still debated. The debate stems in part from the lack of systematic, parametric investigations, as well as from conceptual mistakes in the literature. For instance, demonstrating that a computational algorithm can differentiate between images using their amplitude spectra, or that human observers can detect, or are impaired by, amplitude spectrum manipulations does not mean that observers actually use the amplitude spectrum when both phase and amplitude are available (Gaspar & Rousselet 2009; Wichmann et al., 2010). This conceptual mistake also appears in the brain imaging literature where it takes this form: a difference in brain activity is observed between object categories A and B; a difference is also observed between the phase scrambled versions of these images, leaving the original amplitude spectra intact; it is concluded that the differences observed during the presentation of the unaltered images were due to differences in amplitude spectra. However, this conclusion can only be reached if the original images did not differ in any other aspect than phase and amplitude (for instance color) and if the differences between the original images were abolished after equating amplitude differences.

An example of such problems is found in a recent fMRI study by Andrews, Clarke, Pell, and Hartley (2009). Their results showed larger BOLD responses to faces compared to places in face-preferential brain regions (FFA) for intact images and for their phase-scrambled versions, although the responses were weaker in the latter case. Based on these results, they concluded that at least part of the categorical BOLD differences to intact images could be due to uncontrolled low-level image properties. However, this interpretation is based on a design that included only intact and scrambled images (corresponding to conditions FF, HH, NF, and NH in Figure 1A). Andrews et al. (2009) did not consider critical control conditions in which the intact phase is combined with the amplitude of the opposite category (conditions FH and HF) or with the mean amplitude across categories (conditions Fhf and Hhf in Figure 1A). In the latter case, category-specific variations in the global amplitude spectra are abolished. Thus, if the visual system is sensitive to amplitude spectrum, we should observe different ERP responses towards images with mean versus original amplitude.

Another example of failure to include control conditions with equated amplitude spectra is found in Rossion and Caharel (2011). They reported ERP differences between two categories of color images: faces and cars. The differences were visible as early as 80–100 ms poststimulus onset for both intact and phase
scrambled versions of faces and cars. The authors concluded that the differences observed between the intact picture categories were due to low-level image properties (amplitude spectrum), and not to high-level categorical information. However, this interpretation seems questionable because their early ERP responses could have been driven by differences in contrast or color between the two image categories—a potential confound the authors acknowledged. Moreover, their study did not include necessary control conditions in which amplitude spectra were equated across categories. This conclusion is similar to the one reached by Andrews et al. (2009) for BOLD differences. Importantly, as we already mentioned, ERP research using stimuli controlled for amplitude differences has revealed early categorical differences very similar to those obtained with intact images in those two studies. However, to our knowledge, no study directly compared neural activity to intact and amplitude-equated images; also, no study has systematically assessed ERP
sensitivity to phase and amplitude using parametric designs.

Thus, the objective of this study was to offer the first systematic quantification of the relative contribution of amplitude and phase spectra to the time course of ERPs elicited by faces and objects. Starting from results showing effects of phase information from 100–150 ms onward, our hypotheses were (a) if early categorical ERP differences are due to amplitude spectrum differences between image categories, we should observe early ERP differences to images with different amplitude spectra irrespective of the category (e.g., images of faces, houses, or textures with face amplitude spectra would all differ from images of faces, houses, or textures with house amplitude spectra); (b) if amplitude effects can explain some of the phase effects, sensitivity to amplitude spectrum and its potential interaction with phase should have the same spatial-temporal distributions as phase effects; and (c) similarly, amplitude spectrum sensitivity should have the same direction as phase effects, such that, for instance, if the ERPs were larger for original faces (FF) than for original houses (HH), then at the same electrodes and at the same time points, the ERPs should also be larger to phase scrambled faces (NF) compared to phase scrambled houses (NH). We tested these hypotheses using ERP recordings and images of faces and houses in two experiments. In Experiment 1, we used a 3×3 categorical design as illustrated in Figure 1A. However, because this design confounded categorical and phase effects, we can only conclude about the importance of the amplitude spectrum given a phase/object category. To alleviate this limitation, in Experiment 2 we employed a full parametric design in which we manipulated the Fourier amplitude and phase spectra in 10% intervals, independently for faces and houses. This improved design allowed us to quantify the relative influence of phase and amplitude on brain activity. We analyzed single-subject data using a general linear modelling (GLM) approach. Most subjects were also tested twice to assess the reliability of the effects. Overall, we found a major contribution of phase information on brain activity but no evidence for a contribution of the amplitude spectrum to early ERPs to faces and houses.

Session 2 had the same experimental settings as Session 1, except the sequence of stimuli, which was randomized for each session. Three subjects were members of the lab, including two of the authors, and the remaining five subjects were naive regarding the purpose of the experiment. Two subjects were left-handed, six were right-handed. We tested subjects’ visual acuity using a Colenbrander mixed contrast card set. All subjects had normal or corrected-to-normal visual acuity in the range of 20/25 to 20/10 dec, at 40 cm, 63 cm, and 6 m distances. Subjects’ contrast sensitivity was measured using a Pelli-Robson Contrast Sensitivity Chart, yielding results of 1.95 and above (normal range). All participants also filled in a general health and lifestyle questionnaire. The median number of education years was 21.5 (min = 19, max = 24). All reported good to excellent vision and hearing, and at least weekly exercise. One person reported smoking. All subjects read a study information sheet explaining the behavioral and EEG experimental procedures and provided written informed consent.

**Stimuli**

Two object categories were used in the experiment: faces and houses. There were 10 exemplars of each category (see details in Husk, Bennett, & Sekuler, 2007; Rousselet, Husk, Bennett, & Sekuler, 2008). We manipulated the Fourier amplitude and phase spectra of face and house images using Matlab 2007b (Mathworks, Natick, MA). These manipulations resulted in nine experimental conditions. All images had the same mean pixel intensity and root mean square (RMS) contrast of 0.1 (Figure 1A). The first six conditions were images of faces and houses in which Fourier phase was preserved and the amplitude spectrum was either (a) equated within object category (conditions FF and HH, in which the first letter stands for the phase information and the second letter stands for the amplitude information), (b) swapped between categories (conditions FH and HF), or (c) averaged across categories (conditions Fhf and Hhf). The remaining three conditions, referred to as textures, were created by combining phase spectra of white noise fields with the average amplitude spectrum of faces (condition NF), houses (condition NH), or with the mean amplitude spectrum of faces and houses (condition Nhf). One additional condition, white noise, contained no information at all but is not considered in the present analyses. Figure 1C shows the amplitude spectra of faces, houses, and their average.

**Methods**

**Experiment 1**

**Subjects**

Eight subjects took part in Experiment 1 (six males, two females; median age = 28, min = 20, max = 32). Seven of these eight subjects completed a second session 1 to 3 months after the first one.
asked to position their heads on a chin-rest to maintain a viewing distance of 80 cm and then were given experimental instructions. Participants were asked to discriminate between three image categories: faces, houses, and textures, by pressing one of three keys (1, 2, or 3 on the numerical pad), using three fingers of their dominant hand. The order of the response keys was randomly assigned. Stimuli were presented on a Samsung SyncMaster 1100Mb monitor (Samsung, Samsung Town, Seoul, South Korea) (600 × 800 pixels, height and width: 30 × 40 cm, 21° × 27° of visual angle). All images were 256 × 256 pixel (9° × 9° of visual angle) and were displayed on a gray background (RGB 128, 128, 128) with luminance 33 cd/m². There were 10 blocks, each containing 100 trials with 10 trials for each of the 10 conditions. In each block, each of the 10 trials represented different exemplars. The whole experiment consisted of 1,000 trials. In each trial, first a fixation cross appeared for a random interval between 1000 and 1400 ms, followed by a stimulus presented for five frames (i.e., a maximum of 53 ms), and then by a blank screen which remained displayed until subjects’ response (Figure 1B). The whole experiment lasted about 60 min.

**Behavioral data analysis**

Subjects’ responses were considered correct if they answered “face” to conditions FF, FH, and Fhf, “house” to conditions HF, HH, and Hhf, and “texture” to conditions NF, NH, and Nhf. We used IBM SPSS Statistics 19 (IBM Corporation, Armonk, NY) to compute a 3 × 3 repeated measure ANOVA on the proportion of correct responses. Additionally, we calculated the Harrell-Davis estimator of the median of correct responses across subjects along with percentile bootstrap 95% confidence intervals.

**EEG recording**

EEG was recorded at 512 Hz using the Active Electrode Amplifier System (BIOSEMI [BioSemi B.V., Amsterdam, the Netherlands]) with 128 electrodes mounted on an elastic cap. Two additional electrodes were placed at the outer canthi of the eyes and two below the eyes. During analog to digital conversion, a fifth order Bessel filter was applied to prevent aliasing. The filter had a −3 dB point at one fifth of the sample rate, i.e., 102.4 Hz. Direct coupled data were saved to file.

**EEG data preprocessing**

EEG data were preprocessed using Matlab and the open-source toolbox EEGLAB (Delorme & Makeig, 2004; Delorme et al., 2011). Data were first referenced off-line to an average reference, band-pass filtered between 0.5 Hz and 40 Hz using a two-way least square FIR filter (pop_eegfilt function in EEGLAB) and then epoched from −300 to 1200 ms. The use of a noncausal high-pass filter means that the true onsets of some of the effects could be later than the onsets reported in this paper (Acunzo, MacKenzie, & van Rossum, 2012; Rousselet, 2012; VanRullen, 2011; Widman & Schroeger, 2012). Noisy electrodes were then detected by visual inspection and rejected on a subject-by-subject basis (number of rejected electrodes in Experiment 1: median = 5, min = 0, max = 21; Experiment 2: median = 3, min = 0, max = 18). Baseline correction was performed using the average activity between −300 ms until stimulus onset. Subsequently, we used Independent Component Analysis (ICA), as implemented in the infomax algorithm from EEGLAB. If ICA decomposition yielded components representing noisy electrodes (IC with a very focal, nondipole activity restricted to one electrode while the rest of the map around the electrode is flat), the noisy channels were removed and the ICA was repeated. Components representing blinks, lateral eye-movements or muscle contraction were rejected individually for each subject (number of rejected components in Experiment 1: median = 2, min = 1, max = 5; Experiment 2: median = 2, min = 1, max = 6). After rejection of artifactual components, data were re-epoched between −300 and 500 ms and baseline correction was performed again. Finally, data epochs were removed based on an absolute threshold value larger than 100 μV and the presence of a linear trend with an absolute slope larger than 75 μV per epoch and $R^2$ larger than 0.3. The median number of trials accepted for analysis in Experiment 1 was 989 out of 1,000 (min = 939, max = 995) and in Experiment 2 it was 1,729 out of 1,760 (min = 1,572, max = 1,757).

**EEG data analysis**

Statistical analysis was performed using Matlab 2010a. To determine the ERP sensitivity to phase and amplitude spectra across trials, we performed a 3 × 3 ANOVA for independent groups, with factors φ, amp, and their interaction. The ANOVA was performed at all time points and all electrodes independently using the anovan function from the Matlab Statistics toolbox. We used a bootstrap spatial-temporal clustering approach to correct for multiple comparisons (Pernet, Chauveau, Gaspar, & Rousselet, 2011; Rousselet et al., 2011). In the bootstrap procedure, we mean centered the data independently in each condition, so that the null hypothesis H0 was true (no difference between conditions). The centered data were then sampled with replacement 1,000 times (bootstrap sampling) and a maximum cluster sum of $F$ values was computed for each bootstrap sample using a cluster forming threshold $p = 0.05$. The original $F$ cluster sums were evaluated
against this bootstrapped distribution of maximum cluster sums under H0, considering clusters with \( p \) values < 0.05 (corrected for multiple comparisons) as significant. The reliability of the effects was assessed by comparing the median latencies of the maximum effects between sessions. Post-hoc pairwise comparisons were performed using \( t \)-tests and corrected for multiple comparisons using bootstrap cluster tests. To explore the main effect of phase, we pooled together the ERPs across amplitude levels, separately for faces, houses, and textures. For the main amplitude effect, we pooled together the ERPs across phase levels, separately for face, house, and average amplitude spectra. For each main, we then contrasted: Face versus House, Face versus Texture, and House versus Texture. To explore phase \( \times \) amplitude interactions, we computed a full set of pairwise comparisons between ERPs for individual conditions. Interactions were analyzed only for subjects with significant interaction effects occurring before 200 ms post-stimulus.

**Experiment 2**

**Subjects**

Experiment 2 was carried out 10 to 12 months after Experiment 1. The same eight subjects from Experiment 1 took part in Experiment 2. Six of them completed a second experimental session between 5 and 8 months after the first one. Session 2 was identical to Session 1 except for the order of the stimuli, which was randomized.

**Stimuli**

One face and one house exemplars were randomly chosen from the set of 10 faces and 10 houses used in Experiment 1. The same house and face were used for every subject. We manipulated phase and amplitude spectra of face and house images parametrically and independently (Figure 2). Global phase coherence was altered in 10% intervals, resulting in images containing from 70% to 0% phase coherence (Philastides & Sajda, 2006; Rousselet et al., 2008). Amplitude spectra were manipulated by replacing the original image’s amplitude spectrum with composite amplitude spectra, ranging from 100% face amplitude (0% house amplitude), through 50% face amplitude (50% house amplitude), to 0% face amplitude (100% house amplitude), in 10% steps. These manipulations resulted in 88 face and 88 house conditions, with 8 phase (0%–70%) and 11 amplitude levels (0%–100%). We used a maximum phase coherence level of 70% to reduce the length of the experiment and because increasing phase coherence beyond 70% does not lead to significant behavioral and ERP changes in most subjects (Rousselet, et al., 2008; Rousselet et al., 2009).

**Experimental procedure**

Subjects sat in a booth similar to the one in Experiment 1. Viewing distance, monitor specifications as well as stimulus size and display background color were identical to those of Experiment 1. Subjects were asked to categorize images as faces, houses, or textures by pressing a corresponding key on a computer keyboard. There were 10 blocks, each with 176 trials: 88 images of faces and 88 images of houses. Each of the 88 images represented one of the 88 conditions. There were thus a total of 10 trials per image, category, and condition. The whole experiment contained 1,760 trials and lasted about 75 min, excluding electrode application.

**Behavioral data analysis**

We used Generalized Estimating Equations (GEE, IBM SPSS Statistics 19) to build binary logistic models of the occurrence of “face” and “house” responses in all subjects. We separately modeled the occurrence of
“face” responses within the face stimuli matrix (Figure 2, top panel) and “house” responses within the house stimuli matrix (Figure 2, bottom panel), for each session separately. In the first case, “face” responses were coded as “1” and all the other responses were coded as “0,” whereas in the second case all “house” responses were coded as “1” and the remaining responses were coded as “0.” Both models contained phase and amplitude as within-subject continuous predictors (covariates). The within-subject dependencies were assumed to be homogenous (exchangeable correlation matrix option in SPSS). Model effects were computed using Wald chi-square statistics. Because subjects responded “house” to face stimuli or “face” to house stimuli in only 1%–5% of trials, depending on the session, we did not model this behavior. Statistical analysis of rare events (appearing on less than 10% of the trials) may lead to spurious results.

**EEG recording and pre-processing were as in Experiment 1. EEG data analysis**

Data from individual subjects were analyzed using the LIMO EEG toolbox, a plug-in to the EEGLAB environment (Pernet et al., 2011). Independently at each time point and at each electrode, single-trial ERPs were modeled as:

\[
\text{ERP} = \beta_1 F + \beta_2 H + \beta_3 F \text{amp} + \beta_4 F \varphi + \beta_5 F \text{int} + \beta_6 H \text{amp} + \beta_7 H \varphi + \beta_8 H \text{int} + \beta_0 + \varepsilon \quad (1)
\]

In this model, images of faces (F) and houses (H) were two categorical predictors, whereas global phase coherence (\(\varphi\)), amplitude spectrum (amp), and the phase x amplitude interaction (int) were continuous predictors. Amplitude was coded as the proportion of phase x amplitude interaction (int) were continuous coherence (\(\text{amp}\)). We determined how much unique variance each predictor explained at the max \(R^2\) electrode in each subject by computing semipartial correlation coefficients. Separately for each session and each subject, we measured the maximum unique variance for each predictor in three time windows: P1 (80–120 ms), N1 (130–200 ms), and P2 (200–300 ms). These time windows were defined based on conventions in the ERP literature as well as visual inspection of the data. We used a percentile bootstrap test to determine if, across subjects, the medians of the unique variances differed significantly between phase and amplitude for faces and houses. Medians were estimated using the Harrell-Davis estimator of the 0.5 quantile (Wilcox, 2005).

**Categorical interaction analysis**

If the amplitude spectrum carries categorical information, categorical effects should be modulated by amplitude—in other words, we should observe a category \(\times\) amplitude interaction. We designed a second model to test this interaction (Figure 3, right panel):

\[
\text{ERP} = \beta_0 + \beta_1 F + \beta_2 H + \beta_3 \text{amp} + \beta_4 \varphi + \beta_5 \text{cat} \times \text{amp} + \beta_6 \text{cat} \times \varphi + \varepsilon \quad (2)
\]

In this model we had two categorical variables: faces (F) and houses (H), and four continuous variables: amplitude (amp), phase (\(\varphi\)), category \(\times\) amplitude interaction (cat \(\times\) amp), and category \(\times\) phase (cat \(\times\) \(\varphi\)) interaction.

**Cross-session reliability analysis**

We assessed the reliability of phase and amplitude effects across two experimental sessions by calculating the difference between beta-coefficients obtained from each session at the max \(R^2\) electrode. We tested the significance of the beta differences using a max
temporal cluster bootstrap statistics. First, we pooled together single-trial data from the two sessions. Then we drew two bootstrap samples, with replacement, from the pooled data. The number of trials in each bootstrap sample was the same as the number of trials originally recorded in each session. The bootstrapped samples for each session were then analyzed using the corresponding GLM. Bootstrap sampling and model fitting was performed 1,000 times. Using these 1,000 iterations, we calculated percentile bootstrap univariate 95% confidence intervals for the beta coefficient differences across sessions. Next, the confidence intervals were used to define temporal clusters of absolute beta differences in the bootstrapped and the original data. Original cluster sums of absolute differences were considered significant if they were larger than the 95th percentile of the corresponding bootstrap distribution of maximum cluster sums (i.e., $p < 0.05$ corrected for multiple comparisons).

### Results

#### Experiment 1

**Behavior**

Subjects performed the categorization task very well. Because there were 100 trials per condition, the number of correct responses in Table 1 directly corresponds to percent correct per condition. Participants’ median numbers of percent correct responses in the nine conditions were between 96% and 99% in both sessions (Supplementary Table 1). Subjects’ accuracy was not significantly modulated by phase, Session 1: $F(1.03, 7.12) = 0.991, p = 0.355$; Session 2: $F(1.03, 6.18) = 0.353, p = 0.580$; amplitude, Session 1: $F(1.52, 10.64) = 1.493, p = 0.258$; Session 2: $F(1.48, 8.93) = 0.961, p = 0.392$; or by a phase $\times$ amplitude interaction, Session 1: $F(2.14, 15.02) = 0.294, p = 0.764$; Session 2: $F(2.03, 12.20) = 1.270, p = 0.316$. Due to sphericity violation, all $F$-test values are after Greenhouse-Geisser correction.

**EEG results**

Experiment 1 investigated ERP sensitivity to image phase and amplitude spectra and their interaction ($3 \times 3$ ANOVA). ERPs to faces, houses, and textures replicated a typical pattern of the P1, N1, and P2 components found, for instance, in Rousselet et al. 2008 (Figure 6). The ERPs of all subjects showed a significant effect of category/phase information, a result also reliable across sessions (Figure 4; Supplementary Figures S1 and S2). The median latency of the maximum phase sensitivity was 162 ms in Session 1 (95% confidence interval = 155, 169) and 165 ms (153, 172) in Session 2 (difference = −4 ms [−11, 7]). In all subjects, early phase effects (<200 ms) had lateral-occipital scalp distributions. Pairwise comparisons revealed that, in the N1 time window, ERPs to faces were significantly more negative compared to ERPs to houses and textures; in turn, ERPs to houses were significantly more negative than ERPs to textures (Figure 4, panel B).

Within the first 200 ms poststimulus, significant sensitivity to amplitude spectrum occurred in four subjects in Session 1 (GAR, KWI, CXM, and BTM) and in five subjects in Session 2 (GAR, MAG, WJW, CXM, and BTM; Figure 5, Supplementary Figures S1 and S3). Only three subjects showed significant amplitude effects reliably across sessions (GAR, CXM, and BTM). The median peak latency of the early amplitude sensitivity was 142 ms (127, 172) in Session 1 and 143 ms (125, 178) in Session 2 (difference = −1 ms [−6, 8]). Contrary to the lateral-occipital scalp distribution of phase effects, the early (<200 ms) amplitude effects were observed at medial occipital electrodes. At these electrodes, pairwise comparisons revealed that the only ERP difference that was present in five subjects was such that ERPs to images with face amplitude were significantly more positive than ERPs to house amplitude images (Figure 5, panel B). However, this result is opposite to the ERP pattern observed for the phase effect, in which ERPs to faces were more negative or negative going than ERPs to houses. Five subjects showed also later amplitude effects after 200 ms (KWI, MAG, TAK, WJW, and CXM; Supplementary Figures S1 and S3).

Phase $\times$ amplitude interactions were present in all except two subjects (MAG and CXM) in Session 1 and in all except one subject (CXM) in Session 2. These interactions reached a maximum after 200 ms (Supplementary Figures S1 and S4). Scalp topographies of the interaction effects were mostly left or right occipital. However, the timing of the strongest phase $\times$ amplitude interactions varied substantially among

<table>
<thead>
<tr>
<th>Peak latency S1 – S2 difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faces</td>
</tr>
</tbody>
</table>

Table 1. Median latencies (ms) of peak sensitivities to phase and amplitude spectra. Notes: The two right-hand columns contain differences between sessions for each effect. Square brackets contain 95% percentile bootstrap confidence intervals.

### Notes

The two right-hand columns contain differences between sessions for each effect. Square brackets contain 95% percentile bootstrap confidence intervals.
Figure 4. (A) ERPs to faces, houses, and textures. ERPs were collapsed across amplitude levels separately for faces, houses, and textures and for each session. Each row shows data for one subject and the last row shows grand averages across subjects. Black horizontal lines indicate when the main phase effect was significant. (B) ERP differences. Horizontal lines indicate time windows with significant face–house differences in red, face–texture differences in blue, and house–texture differences in green. The electrode with maximum phase effect, at which the data were plotted, is indicated in the top left corner.
Figure 5. (A) ERPs to face, house, and mean amplitude spectrum images. ERPs were collapsed across phase levels separately for faces, houses, and textures for both sessions. Each row shows data for one subject and the last row shows grand averages across subjects. The black horizontal line indicates when the main amplitude effect was significant. (B) ERP differences. Horizontal lines indicate significant differences: face–house amplitude in red, face–mean amplitude in blue, and house–mean amplitude in green. The top left corner of each plot contains the electrode with maximum amplitude effect at which the data is plotted.
subjects and between sessions: from 240–260 ms for KWI, MAG, BTM, and CMG, to 400–440 ms for TAK or WJW. Three participants (GAR, MAG, and BTM) showed also weak interaction effects around 180 ms poststimulus but these effects were visible in only one session of each participant. Pairwise comparisons at the electrode with maximum interaction effect revealed that these early interactions were explained by more negative ERPs to original house images than to original face images, but more positive ERPs to houses with face amplitude than to faces with house amplitude. For the only three subjects (GAR, MAG, and BTM) showing an effect of amplitude spectrum and an interaction with the category before 200 ms, the face-house differences followed an unexpected pattern, being associated with the larger N170 in the swapped amplitude condition, intermediate for the normalized condition, and the smallest for regular images. Moreover, face-house ERP differences were significant only for one or two amplitude conditions.

To summarize, we hypothesized that (a) early ERP differences between images with different amplitude spectra (FF vs. HH) should be independent of the phase content if they were only due to amplitude spectrum differences (FH vs. HF); (b) sensitivity to amplitude and its potential interaction with phase should have the same spatial-temporal distribution as phase effects; and (c) amplitude effects should have the same direction as phase effects. Our results showed that phase effects were the strongest in all subjects, peaking at lateral-occipital electrodes around 160–170 ms. In contrast to phase, amplitude spectrum effects were much weaker and less reliable across sessions and had a different (medial occipital) scalp distribution. They were also going in a direction opposite from the direction observed for phase effects. Similarly, phase × amplitude interaction effects were weak, observed only in few subjects and varied substantially in timing (predominantly present beyond 200 ms) and topography across participants.

In conclusion, the strength and consistency of ERP phase sensitivity suggests that the phase spectrum is the main contributor to ERP variance as measured with our single-trial analysis. However, amplitude and phase × amplitude interaction effects are difficult to interpret due to their intersubject and intersession variability, which in turn could be due to a lack of sensitivity in our categorical experimental design. For that reason, in Experiment 2 we introduced a parametric manipulation of amplitude and phase.

**Experiment 2**

**Behavior**

The behavioral data indicated that subjects relied more on phase than on amplitude spectrum to categorize images (Figure 6). There was a significant effect of phase on the number of subjects’ face responses to face stimuli and on the number of house responses to house stimuli in both sessions, Faces Session 1: Wald Chi-Square (WCS) = 12.160, p < 0.001; Faces Session 2: WCS = 268.622, p < 0.001; Houses Session 1: WCS = 24.044, p < 0.001; Houses Session 2: WCS = 11.148, p = 0.001. Participants’ accuracy significantly improved as global phase coherence of images increased from 0% to 70%, Faces Session 1: β = 11.300, 95% Wald Confidence Interval = (4.9, 17.6); Faces Session 2: β = 18.438 (16.2, 20.6); Houses Session 1: β = 9.581 (5.752, 13.411); Houses Session 2: β = 12.148 (5.017, 19.279). Significant effects of amplitude on subjects’ face responses to face stimuli or house responses to house stimuli were present only in Session 2 (faces: WCS = 8.446, p = 0.004; houses: WCS = 4.289, p = 0.038), but not in Session 1 (faces: WCS = 3.219 p = 0.073; houses: WCS = 1.715, p = 0.190). In Session 2, the beta coefficient associated with a significant amplitude effect for face stimuli was positive, β = 1.514 (0.493, 2.536), indicating that participants responded “face” more often when the amplitude changed from 0% to 100% face. In the same session, the beta coefficient for house stimuli was negative, β = −3.989 (−7.763, −0.214) suggesting that subjects pressed “house” less often when the amplitude increased from 0% to 100% face. There was no interaction between phase and amplitude in any of the sessions. In sum, the behavioral results suggest that phase is the main contributor to subjects’ categorization decision. However, participants’ responses appear to be influenced also by the congruency between phase and amplitude spectra—if a stimulus contained amplitude and phase information of a face, the number of “face” responses increased; if a stimulus had house phase spectrum and face amplitude, the number of “house” responses decreased. However, this result was not consistent across sessions; hence it is difficult to interpret.

**EEG results**

The results from Experiment 2 confirmed and extended the results from Experiment 1, suggesting again that phase is the main contributor to early categorical ERPs in humans. The max $R^2$ electrode in each participant was also the electrode showing maximum phase sensitivity. Phase effects peaked at around 170 ms after stimulus onset for faces, and slightly later, at 185 ms, for houses (Table 1); at these latencies, amplitude spectrum and phase × amplitude spectrum interaction effects were negligible (Figure 7, Supplementary Figures S5 through S10). Phase effects before 200 ms poststimulus were visible at lateral-occipital electrodes in all subjects (Figure 8,
Supplementary Figures S5 and S6). In the six subjects tested twice, the time-courses of the beta coefficients associated with phase effects were also reliable, although significant differences in latency, magnitude, or both were observed in three subjects (MAG, GAR, and KWI; Supplementary Figures S15 and S16).

Large amplitude spectrum effects occurred before 200 ms after stimulus onset at medial occipital electrodes in three out of eight subjects for faces and in six out of eight subjects for houses (Figure 8). The medial occipital amplitude sensitivity peaked at the median latency of 160 ms (session 1) and 179 ms (session 2) for faces, and at 144 ms (session 1) and 176 ms (session 2) for houses (Table 1) with no significant differences between sessions (Table 1). The latencies of maximum early sensitivities to phase and amplitude differed only for houses in session 1 (Table 2)—the strongest amplitude effects occurred significantly earlier than phase effects. Additionally, some subjects showed also amplitude effects after 200 ms poststimulus, at lateral-occipital electrodes.

Finally, only three subjects had significant interaction effects, and they occurred after 200 ms poststimulus: one subject (KWI) showed a significant effect.
Figure 7. Time course of effects associated with each predictor in Experiment 2. Row 1: mean $F$ values at max R2 electrode; row 2: mean of max $F$ values across all electrodes (envelope); row 3: mean beta coefficients at max R2 electrode; row 4: mean unique variance at max R2 electrode; row 5: mean of the max unique variance across electrodes. Shading represents 95% confidence intervals around the means.
interaction for faces in both sessions and in one session for houses (Supplementary Figures S9 and S10); two subjects had an interaction effect in one session for face stimuli.

**Direction of the effects:** Beyond the mere presence or absence of phase and amplitude effects, it is also important to consider the direction of the effects, as indicated by the sign of the beta coefficients. At the max $R^2$ electrode, between about 130 and 200 ms, the phase beta coefficients were negative in all participants which means that ERPs to faces and houses (i.e., the N170) became more negative as phase changed from 0% (stimuli perceived as textures) to 70% (stimuli perceived as faces or houses—Figure 7, row 3; Figure 9; Supplementary Table 2; Supplementary Figures S11 through S14). In most subjects and most sessions, amplitude spectrum effects had the opposite direction: a positive beta coefficient indicated that as the amplitude spectrum changed from 0% to 100% face amplitude, ERPs to faces and houses became more positive. The same pattern of amplitude effect was observed at the max $R^2$ electrode (Figure 9; Supplementary Figures S11 through S14, columns 2 and 4), and at the electrodes where maximum sensitivity to amplitude was found (Supplementary Figures S11 through 14, columns 3 and 6). The only cases where beta coefficients for phase and amplitude went in the same direction appeared in subject WJW (faces in session 1; houses in both sessions) and TAK (houses in session 2), but the similarities were short-lived and mostly present beyond 200 ms poststimulus.

**Unique variance:** The phase spectrum had significantly larger unique explained variance than the amplitude spectrum for both faces and houses, in the N1 and P2 time windows, but not in the P1 time window (Figures 10 and 11). This was the case at the max $R^2$ electrode and at the maximum across all electrodes (envelope; Supplementary Table 3). In contrast to phase, the amplitude spectrum explained close to zero unique variance in the N1 time window, at the max $R^2$ electrode (Supplementary Table 4, Figures 10 and 11). However, amplitude did explain small amount of unique variance at medial occipital electrodes in the P1 and N1 time windows and at lateral-occipital sites in the P2 time window (Supplementary Table 4, Figure 12). The phase $\times$ amplitude interaction explained

<table>
<thead>
<tr>
<th>Phase vs. amplitude Session 1</th>
<th>Phase vs. amplitude Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Faces</strong></td>
<td><strong>Houses</strong></td>
</tr>
<tr>
<td>Peak</td>
<td>8 [−6, 15]</td>
</tr>
<tr>
<td></td>
<td>40 [21, 54]</td>
</tr>
</tbody>
</table>

Table 2. Differences between median latencies (ms) of maximum sensitivity to phase and amplitude: phase versus amplitude, separately in faces and houses. Notes: A negative difference means an earlier effect for phase compared to amplitude. Square brackets contain 95% percentile bootstrap confidence intervals.
nearly no unique variance in any time window. Overall, within the first 200 ms poststimulus, phase accounted for the largest part of the ERP variance with a lateral-occipital scalp distribution, whereas amplitude had a weak contribution at medial occipital electrodes.

**Phase and amplitude interactions with image category:** Results of the second model analysis revealed that only phase, but not amplitude, interacted significantly with image category (Figure 13; Supplementary Table 5; Supplementary Figures S17 and S18) indicating that categorical differences are modulated by phase, but not by amplitude. This effect was significant in all participants and sessions, and had a lateral-occipital scalp distribution in both sessions (Figure 14). Despite the presence of main amplitude effects in all subjects (apart from TAK, session 2), the category × amplitude interaction was not significant in any of the eight subjects in Session 1 and in seven out of eight subjects.

Figure 9. Time-courses of beta coefficients associated with each predictor of the first regression model of Experiment 2. The betas are from the electrode of the max R2 for each subject (provided on the right hand side of the top left plot). Rows shaded in gray indicate missing subjects in Session 2.
Figure 10. Boxplots of maximum unique variance explained by each predictor at the max R2 electrode, in three time windows: P1 (80–120 ms), N1 (130–200 ms) and P2 (200–300 ms). The boxplots show distributions across subjects. Session 1 is shown in row 1 and Session 2 in row 3. Rows 2 and 4 depict distributions of differences of unique explained variance between phase and amplitude. Near each boxplot, the median difference is indicated with its 95% percentile bootstrap confidence interval.
in Session 2. Subject TAK, who did not show any main amplitude effect, showed sensitivity to category × amplitude interaction in Session 2 at left frontal electrodes.

In sum, the second experiment showed that phase information is the major contributor to the early categorical ERPs observed at lateral-occipital electrodes. As the amount of phase information in the
image increases, this image is perceived more as a face or a house and the corresponding N1 becomes more negative. In contrast, amplitude spectrum sensitivity is mostly observed at medial-occipital electrodes in the P1-N1 window and as amplitude information becomes more face-like, the corresponding ERPs become increasingly positive. Furthermore, phase spectrum explains significantly more unique ERP variance in the N1 and P2 time windows than amplitude and only phase interacts significantly with image category.

### Discussion

Overall, the results from our two experiments suggest that early ERPs to faces and objects are mostly modulated by the phase spectrum, not by the amplitude spectrum. First, in contrast to phase effects, amplitude effects were very weak, inconsistent across sessions and across subjects. Second, amplitude sensitivity before 200 ms poststimulus occurred at medial-occipital electrodes, rather than at the lateral-occipital electrodes that showed the strongest phase effects and categorical ERP differences. Third, as expected, the early ERPs over lateral-occipital sites were becoming increasingly negative as phase coherence of images increased from 0% (noise) to 70%. In contrast, Fourier amplitude modulated ERPs in the opposite direction: ERPs became increasingly positive as images contained increasing amount of face Fourier amplitude. Fourth, the amplitude spectrum accounted for little unique ERP variance compared to phase. Finally, only phase but not amplitude interacted with categorical differences. Overall, these results suggest that the phase spectrum is the main contributor to early categorical ERP differences, whereas amplitude’s contribution is much weaker and present mostly beyond 200 ms poststimulus. This conclusion seems to be a fair interpretation of our results, and could well apply to a larger range of stimuli. However, because we used a simplified set of cropped faces and houses, future studies will have to investigate if our results hold for more realistic stimuli, too. At least, our study provides a systematic approach to tackle this sort of empirical problems.

Our findings question the claims of Rossion and Caharel (2011) and Andrews et al. (2010) that the amplitude spectrum can be responsible for visual categorical differences similar to those observed with intact images. It seems likely that the effects observed in Rossion and Caharel (2011) were not due to amplitude spectrum differences, but instead due to differences in color between their two image categories. In keeping with this idea, it has been shown that the presence of color that is diagnostic for a given category (e.g., green for forest) can speed up early categorization processes reflected in ERPs (Goffaux et al., 2005).

An alternative, simpler explanation might also account for the results of the fMRI study by Andrews et al. (2010). In that study, they found BOLD differences between phase-scrambled images of faces and houses. Instead of being driven by amplitude spectrum differences, their BOLD differences could have been due to differences in image orientation. Indeed, in their study, intact and scrambled images were both vertical for faces and horizontal for places. Image orientation could have created expectations sufficient to influence BOLD responses in FFA and PPA, as suggested by recent studies that have shown that expecting a face can boost neural responses to noise stimuli in the lateral-occipital-temporal cortex (Hansen, Farivar, Thompson, & Hess, 2008; Smith, Gosselin, & Schyns, 2012).

Additionally, Rossion and Caharel (2011) and Andrews et al. (2010) failed to include important control conditions, making it even more difficult to validate their claims. In our study, we attempted to overcome the above shortcomings by either adding control conditions or using a parametric experimental design that provides a more sensitive way of capturing ERP variability associated with changes in physical image properties. Using a general linear model approach, we were able to show that despite the presence of main amplitude effects in almost all subjects, these effects could be dissociated from phase effects: they differed in timing, strength, scalp distribution, and reliability, and they did not interact with image category. Our results converge with the explanation proposed by Clarke et al. (2012), who suggested that cortical sensitivity to amplitude spectrum arises not because amplitude carries information about image category but because the visual system detects and responds to changes in spatial frequency content of the visual input, regardless of its category. In keeping with this idea, a recent study by Hansen, Johnson, and

### Table 3. Differences between median latencies (ms) of maximum sensitivity to phase and amplitude: faces versus houses, separately for phase and amplitude.

<table>
<thead>
<tr>
<th>Faces vs. houses Session 1</th>
<th>Faces vs. houses Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase</td>
<td>Amplitude</td>
</tr>
</tbody>
</table>

Notes: A negative difference means an earlier effect for faces compared to houses.
Figure 12. Topographic maps of unique explained variance for each predictor: face phase (panel A1), house phase (A2), face amplitude (B1), and house amplitude (B2). Each row shows color-coded maps for one subject at the time point of maximum unique variance in three time windows: 80–120 ms (P1), 130–200 ms (N1), and 200–300 ms (P2). The time point (ms) at which each map is plotted is shown above the map. We used different scales for phase and amplitude, because the amount of unique explained variance for amplitude was small compared to phase.
Ellemberg (2012) shows relatively strong modulations of early ERPs by the spatial frequency content of pictures of natural scenes. This implies that the mere presence of ERP differences between images with different amplitude spectra is not sufficient to conclude that image categorization can be achieved by relying purely on the amplitude spectrum (Rousselet, Pernet, Caldara, & Schyns, 2011b; VanRullen, 2011). Moreover, VanRullen and Thorpe (2001) have demonstrated that even though ERPs at about 100 ms poststimulus can differentiate between visually distinct image categories, only later neural activity, beyond 150 ms, correlates with subjects’ decision (see also Philiastides, Ratcliff, & Sajda, 2006; Philiastides & Sajda, 2006).

Instead of explicit categorical processing, the early (<200 ms) main amplitude effects could reflect the visual system’s sensitivity to distributions of local contrast energy. ERP sensitivity to contrast statistics of images of natural textures has been reported recently: it reached a maximum at the medial occipital electrode Oz in the time-window 100–200 ms (Groen, Ghebreab, Lamme, & Scholte, 2012a; Groen, Ghebreab, Lamme, & Scholte, 2012b; Hansen et al., 2012).
Figure 14. Topographic maps of the frequency of effects from the second regression model of Experiment 2. Maps are color-coded according to the number of subjects showing the effects at a given electrode, from 0 subjects (dark blue) to the maximum number of subjects for that session (dark red). Each row shows one effect between time 80 and 300 ms poststimulus.
This scalp distribution and timing fits very well with the amplitude effects found in our study.

Finally, we would like to stress that our results are not an invitation to and should not be used to justify not controlling low-level stimulus parameters. To the contrary, tight control over physical properties of visual stimuli is absolutely crucial to study higher-level object categorization (Rousselet & Pernet, 2011). As we already mentioned, our results might be limited to the simplified cropped faces and houses we used as stimuli. Also, we did not vary task constraints, and therefore cannot rule out a contribution of amplitude spectrum information in different tasks (Rousselet, Pernet, Caldara, & Schyns, 2011). Also, the amplitude spectrum did have an effect at some electrodes and in some subjects: in different experimental contexts, these ERP differences could be misinterpreted as high-level categorical responses. Thus, we cannot stress enough the need to control image properties and use parametric designs to study object categorization.

Acknowledgments

This research was funded by a Leverhulme Trust Research Project Grant (ref: F/00 179/BD). We would like to thank Dr Christoph Scheepers for his help with Generalized Estimating Equations.

Commercial relationships: none.

Corresponding author: Magdalena Maria Bieniek. Email: Magdalena.Bieniek@glasgow.ac.uk.
Address: Institute of Neuroscience and Psychology, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow, UK.

References


Inverting houses and textures: Investigating the characteristics of learned inversion effects. Vision Research, 47(27), 3350–3359.


Rousselet, G. A., & Pernet, C. R. (2011). Quantifying the time course of visual object processing using


