

# The contribution of amplitude and phase spectra-defined scene statistics to the masking of rapid scene categorization

Bruce C. Hansen

Department of Psychology & Neuroscience Program,  
Colgate University, Hamilton, NY, USA



Lester C. Loschky

Department of Psychology, Kansas State University,  
Manhattan, KS, USA



Viewers can recognize the gist of a scene (i.e., its holistic semantic representation, such as its category) in less time than a single fixation, and backward masking has traditionally been employed as a means to determine that time course. The masks used in those paradigms are often characterized by either specific amplitude spectra only, or amplitude and phase spectra-defined structural properties. However, it remains unclear whether there would be a differential contribution of amplitude only or amplitude + phase defined image statistics to the effective backward masking of rapid scene categorization. The current study addresses this issue. Experiments 1–3 explored amplitude spectra defined contributions to category masking and revealed that the slope of the amplitude spectrum was more important for modulating scene category masking strength than amplitude orientation. Further, the masking effects followed an “amplitude spectrum slope similarity principle” whereby the more similar the amplitude spectrum slope of the mask was to the target’s amplitude spectrum slope, the stronger the masking. Experiment 5 showed that, when holding mask amplitude spectrum slope approximately constant, both categorically specific unrecognizable amplitude only and amplitude + phase statistical regularities disrupted rapid scene categorization. Specifically, the masking effects observed in Experiment 5 followed a *target-mask categorical dissimilarity principle* whereby the more dissimilar the mask category is to the target image category, the stronger the masking. Overall, the results support the notion that amplitude only or amplitude + phase-defined image statistics differentially contribute to the effective backward masking of rapid scene gist recognition.

(Carter & Henning, 1971; K. K. De Valois & Switkes, 1983; Henning, Hertz, & Hinton, 1981; Legge & Foley, 1980; Losada & Mullen, 1995; Sekuler, 1965; Solomon, 2000; Stromeyer & Julesz, 1972; Wilson, McFarlane, & Phillips, 1983), and has recently been argued to also serve as an effective means to investigate the time course and spatial mechanisms involved in scene gist recognition, which has typically been operationalized in terms of rapid scene categorization (e.g., Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005; Fei-Fei, Iyer, Koch, & Perona, 2007; Guyader, Chauvin, Peyrin, Hérault, & Marendaz, 2004; Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Kaping, Tzvetanov, & Treue, 2007; Loschky, Hansen, Sethi, & Pydimari, 2010; Loschky & Larson, 2008; Loschky et al., 2007, see also Wichmann, Braun, & Gegenfurtner, 2006, for animal detection). Rapid scene categorization masking paradigms typically employ backward masking (though some paradigms involve the direct manipulation of target scenes and can be construed as rudimentary forms of simultaneous masking).

Backward masking paradigms often utilize masks that are characterized by specific second and higher order statistical regularities of luminance contrast within different real-world scene categories (i.e., regularities that are captured by the Fourier amplitude spectrum or phase spectrum, respectively). A detailed account of these relationships is provided in the Supplementary material. Briefly, the masks typically consist of spatial contrast characteristics defined only by amplitude spectral relationships (e.g., phase-scrambled scenes with amplitude spectra that vary in the distribution of contrast across spatial frequency or orientation, hereafter referred to as *amplitude-only masks*), or some combination of both amplitude and phase spectral relationships (e.g., masks where the amplitude spectra properties are held constant, but

## Visual masking of rapid scene categorization

The use of visual masking has a long history of illuminating the mechanisms of early spatial vision

Citation: Hansen, B. C., & Loschky, L. C. (2013). The contribution of amplitude and phase spectra-defined scene statistics to the masking of rapid scene categorization. *Journal of Vision*, 13(13):21, 1–21, <http://www.journalofvision.org/content/13/13/21>, doi:10.1167/13.13.21.

with the statistical relationships in the phase spectra being systematically varied while preserving various image features, hereafter referred to as *amplitude + phase masks*). Specifically, amplitude + phase masks are designed to possess the lines and edges in scenes that often yield recognizable image structure. Since one needs some modicum of local contrast to see phase-defined image features, we refer to such masks as amplitude + phase defined (but note that the critical manipulations are often focused on manipulations made to the phase spectra).

Most scene gist studies that employ backward masking to investigate the spatial or temporal aspects of gist processing use either amplitude-only masks or amplitude + phase masks. Thus, comparing masking results from those studies raises questions about whether markedly different masking functions would be observed depending on whether amplitude-only masks or amplitude + phase masks were used. This issue is compounded by the fact that few studies have examined the extent to which specific amplitude-only or amplitude + phase mask spectral characteristics contribute to the resulting masking functions. Thus, there are several gaps in our knowledge regarding (a) whether the two types of masks will produce different masking functions, and if so (b) how the different mask spectral characteristics (e.g., amplitude-only or amplitude + phase-defined) modulate the masking functions. Thus, the current study addresses these questions regarding backward masking. We have focused on backward masking because, in contrast to simultaneous masking, it ensures that the participants will have direct access to all target image information (in its original unmanipulated state) prior to being exposed to a mask. Additionally, by varying the stimulus onset asynchrony (SOA), backward masking allows the experimenter to assess when amplitude or phase information becomes relevant to the categorization process (but see VanRullen, 2011), though here, we primarily focus on the spatial aspects of the masking. Below, we provide a more detailed account of how research on backward masking has raised these crucial theoretical questions regarding amplitude-only masks and amplitude + phase masks.

### Amplitude-only backward masking

Amplitude-only backward masking effects have been investigated in a pair of studies by Loschky et al. (2007) and Loschky et al. (2010), which examined masking effects on rapid scene categorization performance. Those studies used masks created by phase-randomizing scenes, thereby leaving only amplitude information intact, which we therefore refer to as amplitude-only masks. Those masks were unrecognizable, and thus

devoid of any semantic meaning, which eliminated masking effects caused by conceptual processes, known as *conceptual masking* (e.g., Intraub, 1984; Loftus & Ginn, 1984; Loftus, Hanna, & Lester, 1988; Michaels & Turvey, 1979; Potter, 1976). Doing so crucially limited masking effects to those due to visual mechanisms responding to image structure defined by amplitude spectral properties, rather than those attributable to conceptual masking. Loschky et al. (2007) and Loschky et al. (2010) specifically examined masking effects produced by three types of amplitude-only masks:

- (a) phase-randomized versions of scenes from different categories than the target scenes (i.e., the masks had different amplitude spectral characteristics than the target scenes),
- (b) phase-randomized versions of the target images themselves (i.e., the masks had identical amplitude spectral characteristics to the target scenes), or
- (c) white noise masks (i.e., masks lacking any global amplitude- or phase-defined structural characteristics—see Supplementary material for further detail).

These comparisons produced two critical results: (a) there were no differences in the masking between amplitude-only masks derived from the same versus different scene category than the target, and (b) both masking conditions based on phase-randomized scene masks produced significantly stronger effects than the white noise masking condition. The former result suggests that amplitude spectral properties produce little if any category-specific masking effects. On the other hand, the latter result suggests that general spatial frequency and orientation differences (relative to white noise) play a critical role in masking rapid scene categorization. That is, greater masking was caused by the phase-randomized amplitude-only masks that possessed typical amplitude spectrum characteristics of real-world scenes (i.e., contrast fall-off with increasing spatial frequency, and contrast biases at different orientations), rather than the white noise masks, which did not. More specifically, since the phase-randomized scenes as masks still had intact amplitude spectra, they possessed amplitude features common to most natural scenes including (a) the  $1/f^\alpha$  fall off in contrast with increasing spatial frequency (i.e., the amplitude spectrum slope,  $\alpha$ , typically  $\approx 1.0$ ) and (b) the vertical/horizontal (i.e., cardinal) orientation bias found across all scenes (cardinal > oblique) (see Hansen, Haun, & Essock, 2008, for a review). Thus, the question is which aspect of the amplitude spectrum is responsible for such masking? Specifically, is it (a) the slope,  $\alpha$ , of the amplitude spectrum, (b) the orientation biases, or (c) both, that produced the strongest masking effects for phase-randomized scene masks (compared to white noise) in the Loschky et al.

(2007) and Loschky et al. (2010) studies? Experiments 1–3 aimed at answering this question.

### Amplitude + phase backward masking

The studies by Loschky et al. (2007) found that actual unaltered scenes used as masks (i.e., a type of amplitude + phase mask) were much more effective at interfering with rapid scene categorization than phase-randomized scene masks (i.e., amplitude-only masks). This finding suggests that amplitude-only masks and amplitude + phase masks produce different masking functions. However, the amplitude + phase masks utilized by Loschky et al. (2007) contained recognizable scene structure. Thus, an alternative explanation for such differences in masking effects would be conceptual masking. Loschky et al. (2010) therefore explored this conceptual masking account in a follow-up study using unrecognizable amplitude + phase masks. Those masks were created by feeding real-world scene images to the Portilla and Simoncelli (2000) texture synthesis algorithm, which produced scene-texture masks containing unrecognizable conjoint amplitude + phase-defined image structures. Loschky et al. (2010) found that such amplitude + phase masks interfered with rapid scene categorization far more than amplitude-only (phase randomized) masks, and both masked more than white noise. Critically, the unrecognizable amplitude + phase masks produced 74% of the difference in masking found between white noise masks and unaltered scenes used as masks, namely recognizable amplitude + phase masks. That is, 74% of the observed masking effects mentioned above that might have been attributed to conceptual masking could, instead, be explained by unrecognizable amplitude + phase-defined scene structure.

Loschky et al. (2010) showed that amplitude + phase masks, which contained unrecognizable phase-defined image structure, yielded strong masking effects unexplainable by conceptual masking. Nevertheless, we are still left with the question of how such amplitude + phase regularities modulate rapid scene categorization performance. Given that those masks were derived from real-world scene images, the unrecognizable phase-defined image structures in the amplitude + phase masks may have regulated scene categorization in a category-specific manner. If so, to take an example, would unrecognizable amplitude + phase masks derived from beach scenes more strongly mask beach targets than unrecognizable amplitude + phase masks derived from mountain scenes (or vice versa)? Experiments 4–5 aimed at answering this question.

It is important to note that utilizing backward masking paradigms to study rapid scene categorization is not without its limitations. One point worth

reiterating here is that masking has a long history in the study of early visual processes such as the detection and discrimination of sinusoidal gratings or Gabor patterns. Thus, backward masking effects in rapid scene categorization must be considered in the terms of early spatial vision mechanisms assessed by visual masking. Specifically, one must consider whether a given set of backward masking effects can be predicted by the known response characteristics of early visual mechanisms, which could apply equally well to a wide array of visual tasks including rapid scene categorization. For example, if a given backward masking effect can be accounted for by a reduction of the signal-to-noise ratio in early spatial channels, then the information in the mask may simply have been more effective at stopping informative scene information from being passed along to higher levels where scene categorization is thought to occur (e.g., the parahippocampal place area [PPA] or the lateral occipital complex [LOC]). We therefore apply this approach to interpreting the results of the current study. In doing so, we attempt to dissociate those masking effects more likely to be explained by earlier spatial channel interactions from those less likely to be so explained.

### The current study

To return to the primary goals laid out in our introductory paragraph, the current study: (a) provides an affirmative answer to the question of whether unrecognizable amplitude-only masks versus unrecognizable amplitude + phase masks produce different masking functions; and (b) addresses how the spectral characteristics of amplitude-only masks and amplitude + phase masks modulate their respective masking functions. Experiments 1–3 were concerned with the effects of the spectral characteristics of amplitude-only masks. Experiment 1 employed noise masks containing only amplitude-defined structure (i.e., variable slope,  $\alpha$ , and variable orientation bias magnitudes). Experiment 2 followed up Experiment 1 by extending its crucial masking conditions into the temporal domain. Experiment 3 controlled for the perceived contrast (rather than physical contrast) of the noise masks used in Experiments 1 and 2. Experiments 4–5 were concerned with the effects of the spectral characteristics of amplitude-only versus amplitude + phase masks. Experiment 4 assessed the recognizability of image structure in amplitude-only masks versus amplitude + phase masks that would be used in Experiment 5. This was necessary to ensure minimal to no recognizability in order eliminate conceptual masking effects. Experiment 5 then assessed whether unrecognizable amplitude-only versus amplitude + phase masks interfered

with rapid scene categorization in a category-specific manner.

## General method

### Apparatus

For experiments conducted at Kansas State University (Experiments 1, 3, 4, & 5), all stimuli were presented on 19 in. Samsung SyncMaster 957 MB S monitors. These were driven by Optiplex 170L Pentium 4 processors (2.8 GHz), which were equipped with 512 RAM and Intel 82865G Graphics Controller Integrated Chipset graphics cards, which had 8-bit grayscale resolution. Stimuli were displayed using a linearized look-up table, generated by calibrating with a Color-Vision Spyder2 Pro sensor. Maximum luminance output of the display monitors was  $80.6 \text{ cd/m}^2$ , the frame rate was set to 85 Hz, and the resolution was set to  $1024 \times 768$  pixels. Single pixels subtended  $.0369^\circ$  of visual angle (i.e., 2.21 arc min) as viewed from a distance of 53.3 cm using machined chin rests.

For experiments conducted at Colgate University (Experiment 2), all stimuli were presented on a 21 in. Viewsonic (G225fB) monitor. This was driven by a dual core Intel® Xeon® processor (1.60 GHz x2), which was equipped with 4 GB RAM and a 256 MB PCIe x16 ATI FireGL V7200 dual DVI/VGA graphics card, and with 8-bit grayscale resolution. Stimuli were displayed using a linearized look-up table, generated by calibrating with a Color-Vision Spyder3 Pro sensor. Maximum luminance output of the display monitor was set to yield  $80.6 \text{ cd/m}^2$ , the frame rate was set to 85 Hz, and the resolution was set to  $1024 \times 768$  pixels. Single pixels subtended  $0.0363^\circ$  of visual angle (i.e., 2.18 arc min) as viewed from a distance of 58.0 cm using an Applied Science Laboratories (ASL) chin and forehead rest.

### Scene image database

A total of 600 grayscale scene images (selected from the Internet and free of any copyright restrictions) were used in the current study (100 images per scene category) as either *target* stimuli or as images used to construct *mask* stimuli. There were six basic level categories: (a) beaches, (b) forests, (c) airports, (d) streets, (e) home interiors, and (f) store interiors. These could be further categorized as three conjoint superordinate categories: natural outdoor (beaches and forests), man-made outdoor (airports and streets), and man-made indoor (home and store interiors). Each image category set was assembled by cropping a  $512 \times$

512 pixel region from a random location within the larger image.

Each image,  $I(x, y)$ , was normalized to possess a fixed root-mean square (rms) contrast (as defined in Pavel, Sperling, Riedl, & Vanderbeek, 1987) and mean luminance in the spatial (i.e., image) domain using the sequence of functions that can be found in Hansen and Hess (2007). Here, rms was set to 0.126, with the mean grayscale value of all images fixed at 127. The particular choice of rms allowed for a suprathreshold contrast that did not result in pixel grayscale values outside the 0–255 range.

## Experiment 1

The current experiment was designed to determine which aspect of the amplitude spectrum (slope or orientation) in amplitude-only masks is responsible for masking rapid scene categorization performance: (a) the slope of the amplitude spectrum, (b) the orientation biases, or (c) both?

### Method

#### Participants

Forty-seven Kansas State University students (34 female), all naïve to the purpose of the study, participated for course credit after giving their Institutional Review Board-approved informed written consent (with written parental consent also given for those under the age of 18). All observers had normal (or corrected to normal) vision and their ages ranged from 16 to 30 years ( $M = 19.33$ ,  $SD = 2.93$ ).

#### Noise mask creation

All noise masks were constructed in the Fourier domain using MATLAB (version R2008a) including Image Processing (version 6.1) and Signal Processing (version 6.9) toolboxes. Mask construction involved creating a base amplitude spectrum,  $BASE_{AMP}(f, \theta)$ , weighted by an orientation filter. The orientation filter used here consisted of an orientation biased amplitude spectrum,  $ORNT_{AMP}(f, \theta)$ , constructed by averaging the amplitude spectra from all images described in the General method section. The base and orientation biased spectra were created in polar coordinates, thus  $f$  and  $\theta$  represent the spatial frequency and orientation dimensions, respectively. The orientation filter first involved constructing an amplitude spectrum from the average of the entire set of scene category images ( $n = 600$  total images), shifted into polar coordinates, with the DC (i.e., zero frequency) component assigned a

zero. The averaged amplitude spectrum, denoted as  $ORNT_{AMP}(f, \theta)$ , possessed the typical orientation bias observed in natural scenes, namely more contrast along the cardinal axes, vertical and horizontal, than the obliques (e.g., Coppola, Purves, McCoy, & Purves, 1998; Hansen & Essock, 2004; Keil & Cristobal, 2000; Switkes, Mayer, & Sloan, 1978; Torralba & Oliva, 2003; van der Schaaf & van Hateren, 1996). However, it also contained the typical spatial frequency contrast bias in (i.e., much more contrast at lower than higher spatial frequencies). Since we needed  $ORNT_{AMP}(f, \theta)$  to only serve as an orientation filter, we next removed the spatial frequency contrast bias by the following operations. First, we created an orientation averaged vector,  $OA(f_u)$  by averaging the amplitude coefficients across all orientations for each spatial frequency in  $ORNT_{AMP}(f, \theta)$ , stated formally as:

$$OA(f_u) = \frac{1}{v_n} \sum_{v=1 \dots v_n}^{u=1 \dots u_n} ORNT_{AMP}(f_u, \theta_v), \quad (1)$$

where the subscripts  $u$  and  $v$  represent spatial frequency and orientation coordinates, respectively, with  $u_n$  being the total number of spatial frequencies (in cycles per picture, cpp), and  $v_n$  being the total number of orientations. Importantly, due to the nature of the discrete Fourier transform (DFT) (e.g., Hansen & Essock, 2004), the total number of orientations,  $v_n$ , at a given spatial frequency varies monotonically with increasing spatial frequency. Next, to remove the spatial frequency bias from  $ORNT_{AMP}(f, \theta)$ , each orientation vector in  $ORNT_{AMP}(f, \theta)$  was divided by  $OA(f_u)$ . This produced a flat amplitude spectrum (i.e.,  $\alpha = 0.0$ ), but with the orientation bias preserved, which served as the orientation filter used to weight the amplitude spectra of the noise masks in the current experiment. This is formally stated as:

$$O_{filt}(f, \theta) = \frac{ORNT_{AMP}(f, \theta_{v=1 \dots v_n})}{OA(f_u)}. \quad (2)$$

Here,  $f$  in the numerator is without a subscript as the division is applied to all spatial frequencies along each orientation vector within  $ORNT_{AMP}(f, \theta)$ . Finally,  $O_{filt}(f, \theta)$  represents the orientation filter itself.

Next, the base amplitude spectrum,  $BASE_{AMP}(f, \theta)$ , was constructed in a single  $512 \times 512$  polar matrix, with all coordinates assigned the same arbitrary amplitude coefficient (except the location of the DC component, which was assigned a zero). The result is a flat isotropic broadband spectrum (i.e., possessing equal contrast across all spatial frequencies and orientations). Then, the base spectrum exponent could be adjusted by multiplying each spatial frequency's amplitude coefficient by  $f^{-\alpha}$ , with  $\alpha$  representing the amplitude spectrum exponent. For the current exper-

iment,  $\alpha$  was set to either 0.0, 0.5, 1.0, or 1.5 to capture the typical range of  $\alpha$ s observed in scene imagery (e.g., 0.5 to 1.5), and a white noise condition (i.e.,  $\alpha = 0.0$ ).

Then, an orientation bias was applied to a base spectrum having one of the four different  $\alpha$  values by weighting it with  $O_{filt}(f, \theta)$  according to the following operation:

$$MASK_o = \{(BASE_{AMP}[f, \theta] \times [1 - O_{mag}]) + (O_{filt}[f, \theta] \times [O_{mag}])\}. \quad (3)$$

Here,  $O_{mag}$  was a scalar value controlling the orientation bias magnitude, which took on one of five values from zero to one (in steps of 0.25), with zero producing an isotropic amplitude spectrum and one producing a mask amplitude spectrum having the full bias in  $ORNT_{AMP}(f, \theta)$ . Thus, subscript  $\alpha_o$  represents mask amplitude spectra possessing one of the four amplitude spectrum slopes ( $\alpha$ ) and one of the five orientation bias magnitudes ( $o$ ).

Finally, we used a unique phase spectrum for each mask by assigning random values from  $-\pi$  to  $\pi$  to the coordinates of a  $512 \times 512$  polar matrix,  $\Phi_{RAND}(f, \theta)$ , such that the phase spectra were odd-symmetric—i.e., for  $\theta$  angles in the  $[\pi, 2\pi]$  half of polar space,  $\Phi_{RAND}(f, \theta) = -\Phi_{RAND}(f, \theta)$ . The noise patterns were rendered in the spatial domain by taking the inverse Fourier transform of  $MASK_o(f, \theta)$  and a given  $\Phi_{RAND}(f, \theta)$ , with both shifted to Cartesian coordinates prior to the inverse DFT. The rms contrast of all noise masks was fixed at 0.126 in the spatial domain by scaling the power spectrum in the Fourier domain. A total of 100 unique masks was generated for each of the four  $\alpha$  and five orientation bias magnitude conditions. Examples of the 20 mask types are shown in Figure 1A.

### Design and procedure

Experiment 1 used a six alternative forced choice (AFC) scene categorization backward masking paradigm. The experiment had a 4 (Mask  $\alpha$ )  $\times$  5 (Orientation Bias Magnitude) mixed design, with mask  $\alpha$  a between-subjects variable, and orientation bias magnitude a within-subjects variable. In each mask  $\alpha$  condition, each participant saw 60 different target category stimuli (described in the General method section) for each level of mask orientation bias magnitude (10 stimuli randomly selected without replacement from each scene category), for 300 total trials. The order of target category and mask orientation bias magnitude was randomized. Participants were randomly assigned to between-subjects conditions.

A given trial sequence consisted of a 500 ms fixation point (0.3° black circle) at the screen center (with the screen set to mean luminance), followed by a target scene presented for 23.56 ms, followed by a 11.76 ms

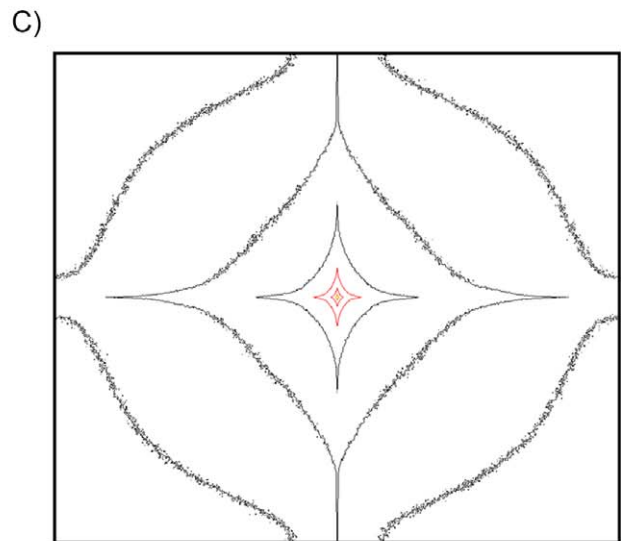
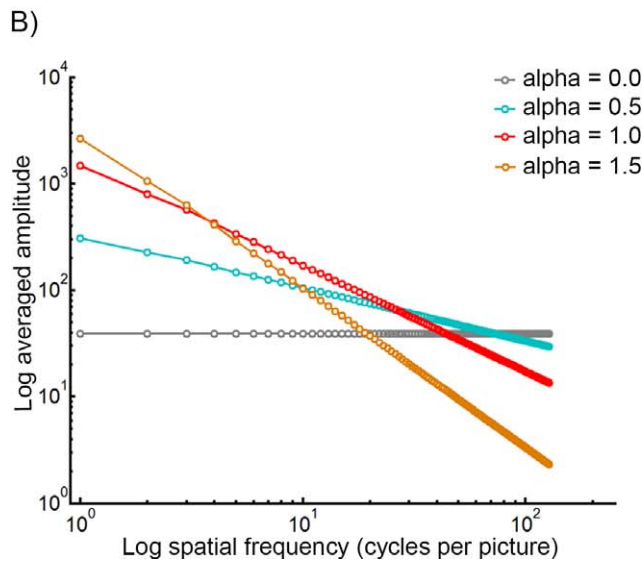
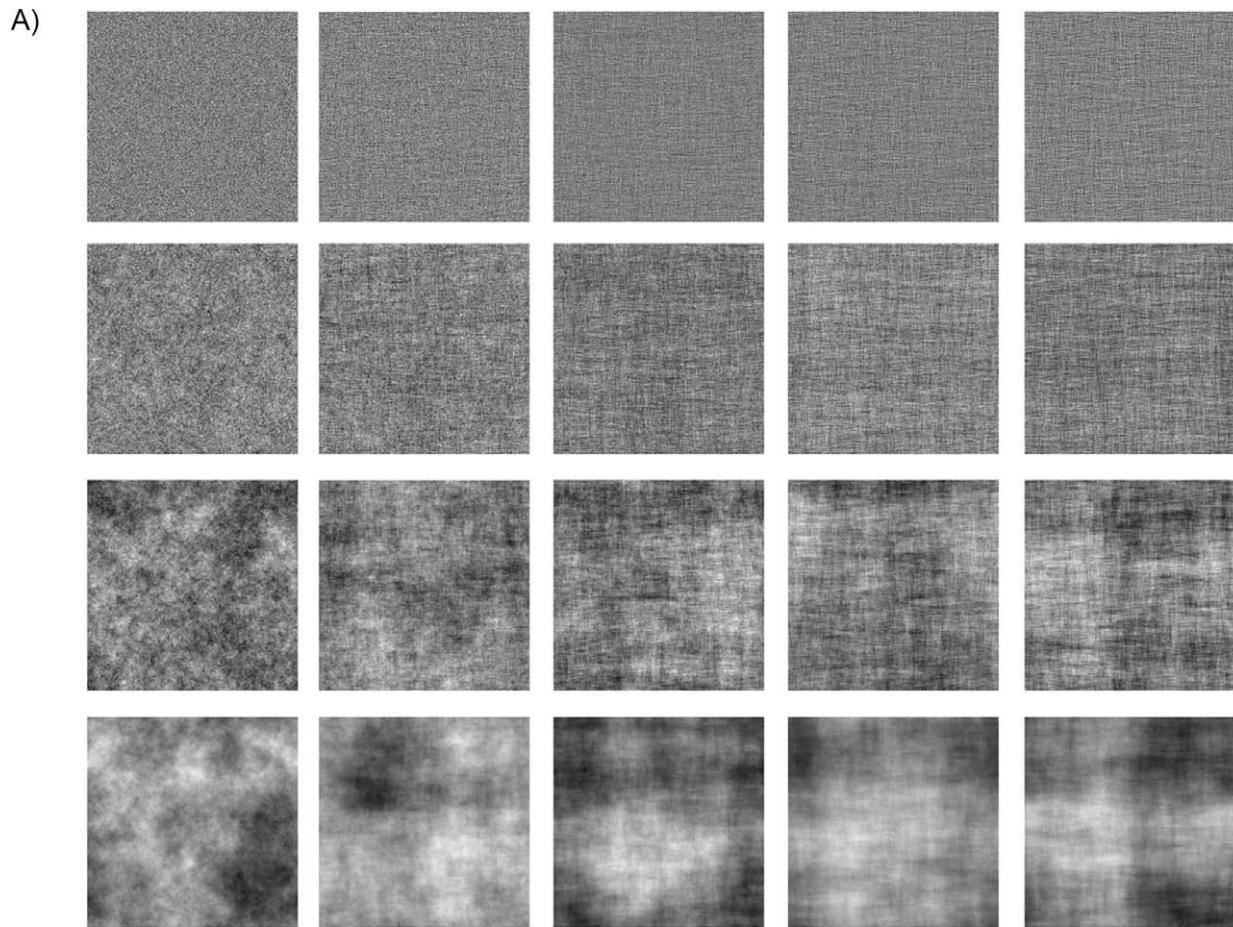


Figure 1. (A) Example noise masks used in Experiment 1. The examples are arranged in rows for amplitude spectrum slope (i.e.,  $\alpha$ ), and columns for orientation magnitude. Specifically, amplitude spectrum slope increases from the top row ( $\alpha = 0.0$ ) to the bottom row ( $\alpha = 1.5$ ) in steps of 0.5. The far left column is  $O_{\text{mag}} = 0.0$  (i.e., no orientation bias) to the far right column,  $O_{\text{mag}} = 1.0$  (maximum orientation bias), in steps of 0.25. (B) Shows orientation averaged amplitude spectra for the four different noise mask slopes. (C) Shows a 2-D contour plot of  $O_{\text{fit}}(f, \theta)$  plotted in polar coordinates. Note the distinct horizontal and vertical orientation bias.

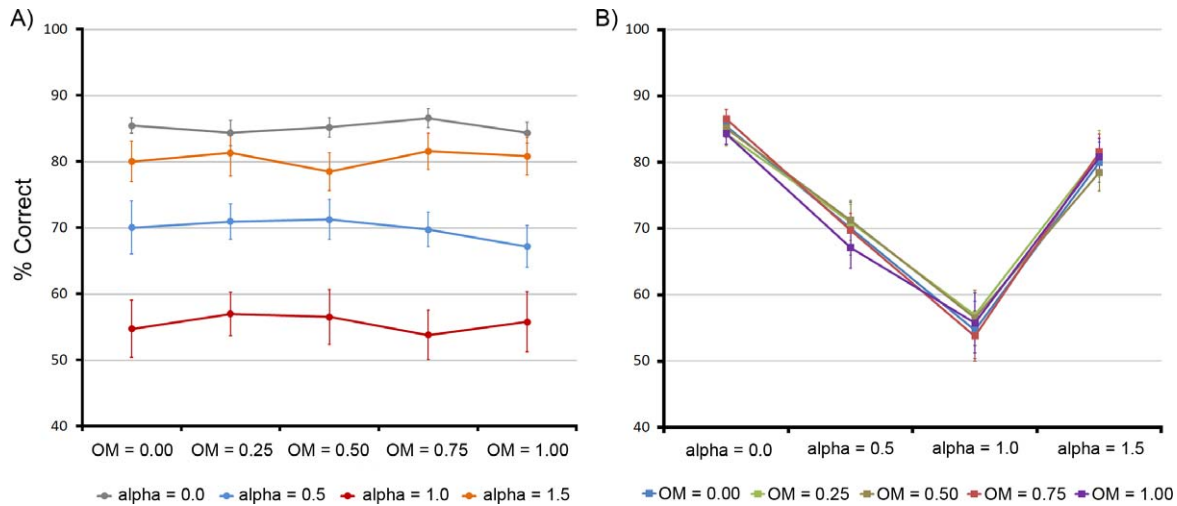


Figure 2. Averaged data for Experiment 1. For both plots, the ordinate shows averaged percent correct (note that the scale begins at 40% to increase visibility). (A) Shows amplitude spectrum slope masking functions for the different mask orientation bias magnitudes (OM). (B) Shows mask orientation bias magnitude (OM) masking functions for the four different mask amplitude spectrum slopes. Error bars are  $\pm 1$  SEM.

interstimulus interval (ISI) (a mean luminance blank screen, stimulus onset asynchrony [SOA] = 35.32 ms), then a noise mask for 70.67 ms (i.e., 3:1 mask:target duration ratio), followed by a mean luminance screen for 750 ms, and then the response screen containing a  $2 \times 3$  grid of the six basic level scene category labels, until the observer responded by a mouse click on the category label matching the target stimulus. The positions of the category labels were randomized on each trial to avoid category location response bias. That is, by randomizing the locations of response choices on every trial, any location-based response bias (e.g., top left response cell) would be randomly distributed across all categories, rather than only a single category. Note however, that this requires participants to first search for the correct answer on each trial, adding to their reaction time, making this method better suited to analyzing accuracy than reaction times. Before the experiment began, participants were given practice trials using target stimuli from categories not used in the experiment, but with noise masks from the assigned  $\alpha$  condition and variable mask orientation bias magnitude.

## Results and discussion

Random assignment of participants to the between-subjects variable, mask amplitude spectrum slope  $\alpha$ , resulted in 12 in the  $\alpha = 0.0$  condition, 11 in the  $\alpha = 0.5$  condition, 11 in the  $\alpha = 1.0$  condition, and 13 in the  $\alpha = 1.5$  condition. Figure 2a shows mean scene categorization accuracy for each amplitude spectrum slope,  $\alpha$ , mask condition as a function of mask orientation bias

magnitude. Figure 2b shows the same data plotted for each level of mask orientation bias as a function of mask amplitude spectrum  $\alpha$ , which we tested with a  $4$  (Mask  $\alpha$ )  $\times$   $5$  (Orientation Bias Magnitude) mixed repeated measures analysis of variance (ANOVA). The data in Figure 2 show a strong masking effect for amplitude spectrum  $\alpha$ , which was confirmed by a significant between-subjects effect of mask amplitude spectrum  $\alpha$ ,  $F(3, 43) = 24.09$ ,  $p < 0.001$ , Cohen's  $f^2 = 0.271$ .<sup>1</sup> On the other hand, Figure 2 shows little to no effect of mask orientation bias magnitude, which is also confirmed by a nonsignificant within-subjects effect of mask orientation bias magnitude,  $F(4, 172) = 0.40$ ,  $p = 0.81$ . Likewise, as suggested by Figure 2, there was no significant interaction between mask  $\alpha$  and mask orientation bias magnitude,  $F(12, 172) = 0.735$ ,  $p = 0.72$ .

We carried out post-hoc comparisons of each amplitude slope condition against the others using a Bonferroni adjusted family-wise  $\alpha$  level of 0.0083 for the six unique comparisons among the four conditions. To carry out the independent groups two-tailed  $t$  tests, we first created equal sized groups by randomly selecting one subject to omit from the  $\alpha = 0$  condition and two subjects to omit from the  $\alpha = 1.5$  condition (though the statistical test results were the same whether the subjects were omitted or not). As suggested by examination of Figure 2a, all three comparisons with the  $\alpha = 1.0$  condition were significant ( $t_s \geq 3.11$ ,  $p_s \leq 0.006$ , Cohen's  $d_s \geq 0.56$ ), as was the comparison of the  $\alpha = 0.5$  and 0.0 (white noise) conditions,  $t(20) = 5.51$ ,  $p < 0.001$ , Cohen's  $d = 1.91$ . Conversely, the comparison of the  $\alpha = 1.5$  and 0.0 conditions was not significant,  $t(20) = 1.63$ ,  $p = 0.118$ , nor was the comparison of the  $\alpha = 1.5$  and 0.5 conditions,  $t(20) = 2.42$ ,  $p = 0.025$  (based on the family-wise adjusted alpha level).

Lastly, while unlikely (see Introduction and Loschky et al., 2007), we examined whether categorization performance for any given scene category was masked more or less compared to the others as a function of mask orientation bias magnitude. There were no significant differences for any of the six scene categories as a function of mask orientation bias magnitude for any of the mask amplitude spectrum slopes,  $p_s > 0.05$  (results not shown). Thus, in terms of which amplitude-only spectral characteristics most strongly mask rapid scene categorization, the current results show that the distribution of contrast across spatial frequency (i.e., amplitude spectrum slope  $\alpha$ ) is much stronger than the distribution of contrast across orientation.

Finally, the masking functions shown in Figure 2b show an interesting trend that seems related to the typical amplitude spectrum slope found in natural scenes (i.e.,  $\alpha \approx 1.0$ ). Specifically, it is tempting to conclude that the more similar the mask slopes were to the typical slope observed in natural scenes, the stronger the masking effect. However, since the entire target image set had a mean slope of 1.12 ( $SD = 0.12$ ), and because there were no category-specific masking effects (across orientation magnitude or slope), Occam's razor suggests the simpler (thus better) description is that the observed masking effects follow an *amplitude slope similarity principle*. Specifically, the more similar the mask amplitude slope was to the target image slope, the stronger the masking effect. We will return to this notion in Experiments 4 and 5, as well as in the General discussion.

## Experiment 2

Given the amplitude spectrum slope was the driving force in the rapid scene categorization masking effects in Experiment 1 for a fixed target/mask SOA, Experiment 2 further explored the crucial conditions from Experiment 1 within the temporal processing domain. Specifically, we investigated whether the amplitude spectrum slope  $\alpha$  was more or less of a factor at varying processing times. Thus, we replicated Experiment 1 for masks set to either  $\alpha = 0.0$  or  $1.0$ , with each possessing either a strong or no orientation bias ( $O_{mag} = 1.0$  vs.  $0.0$ ), across five different target/mask SOAs. We note, along with VanRullen (2011), that the relationship between masking SOA and processing time is not as simple as suggested by many studies using backward masking to study the time course of perception. Specifically, neurophysiological studies that have investigated the link between masking SOA and the time course of target-related brain activity have

shown that while SOA and processing time are not identical, they are related (e.g., Rieger, Braun, Bulthoff, & Gegenfurtner, 2005).

## Method

### Participants

Fifty-three Colgate University students (34 female), all naïve to the purpose of the study, participated for course credit after giving their Institutional Review Board-approved informed written consent. All observers had normal (or corrected to normal) vision, and their ages ranged from 18 to 21 years ( $M = 18.8$ ,  $SD = 0.79$ ).

### Design and procedure

The mask stimuli were identical to those used in Experiment 1, with the following exceptions. Only two mask amplitude spectrum  $\alpha$  values ( $\alpha = 0.0$  and  $\alpha = 1.0$ ) and two mask orientation bias magnitudes ( $O_{mag} = 0.0$  and  $O_{mag} = 1.0$ ) were employed in the current experiment. Target stimuli were drawn from the set used in Experiment 1, as described in the General method section.

The current experiment employed the same paradigm as Experiment 1, except that the current experiment utilized five different SOAs and a no-mask condition. The experiment used a  $2$  (Mask  $\alpha$ )  $\times$   $2$  (Orientation Bias Magnitude)  $\times$   $6$  (SOA & No-Mask) mixed design, with mask  $\alpha$  and orientation bias magnitude being between-subjects variables, and SOA being a within-subjects variable. Within a given mask  $\alpha$  and orientation bias magnitude condition, each participant saw 60 different target stimuli (described in the General method section) for each SOA (with 10 stimuli randomly selected without replacement from each of six scene categories). For the no-mask condition, an additional 60 images were randomly sampled (10 per scene category), for a total of 360 trials. Order of target category and SOA (including no mask) was randomized for each trial. Participants were randomly assigned to the between-subjects conditions.

The trial sequence was identical to that of Experiment 1 except that the target/mask ISI was set to yield five different SOAs (= target duration + ISI): 23.56 ms (i.e., 0 ms ISI), 35.29 ms, 47.05 ms, 70.57 ms, and 117.64 ms. For no-mask trials, the target image was simply followed by the same mean luminance screen for 750 ms as in the masking conditions (as in Experiment 1). Participants' task was the same as Experiment 1, and they were given practice trials using target stimuli from categories not used in the experiment.



## Results and discussion

Random assignment of participants to the between-subjects variable, mask amplitude spectrum slope  $\alpha$  + orientation bias magnitude, resulted in 12 in the  $\alpha = 0.0$  + orientation bias = 0.0 condition, 14 in the  $\alpha = 0.0$  + orientation bias = 1.0 condition, 13 in the  $\alpha = 0.0$  + orientation bias = 0.0 condition, and 14 in the  $\alpha = 1.0$  + orientation bias = 1.0 condition. Figure 3 shows mean scene categorization accuracy for each amplitude spectrum slope,  $\alpha$ , and orientation bias mask condition as a function of SOA (and no mask), and we conducted a 2 (Mask  $\alpha$ )  $\times$  2 (Mask Orientation Bias Magnitude)  $\times$  5 (SOA) mixed repeated measures ANOVA to test the effects of these factors. As expected, Figure 3 shows a reduction of masking strength with increasing SOA for all between-subjects masking conditions, which was confirmed by a significant main effect of SOA,  $F(4, 196) = 135.64, p < 0.001$ , partial  $\eta^2 = 0.74$ . Also, as in Experiment 1, noise masks with amplitude spectrum  $\alpha = 1.0$  produced stronger effects than that of  $\alpha = 0.0$  for the shorter SOAs, as shown by a significant main effect of mask  $\alpha$ ,  $F(1, 49) = 15.71, p < 0.001$ , Cohen's  $f^2 = 0.26$ , and a significant mask  $\alpha \times$  SOA interaction,  $F(4, 196) = 36.45, p < 0.001$ , Cohen's  $f^2 = 1.01$ . Interestingly, the effect of mask orientation bias magnitude was absent in both  $\alpha = 1.0$  conditions, but there was an apparent effect of orientation magnitude in the  $\alpha = 0.0$  conditions. However, this was not supported by the ANOVA, which found neither a significant main effect of orientation magnitude nor any significant interactions involving it ( $ps \geq 0.212$ ). Thus, we explored the significant Mask  $\alpha \times$  SOA interaction to identify the SOAs yielding significant differences between the two mask  $\alpha$  conditions. We collapsed across orientation magnitude within each level of mask  $\alpha$ , and ran independent  $t$  tests (Bonferroni adjusted family-wise  $\alpha$  level = 0.0083) between the two  $\alpha$  groups at each SOA (and the no-mask condition). This showed significant differences between the two mask  $\alpha$  conditions, 24 ms:  $t(50) = 7.89, p < 0.001$ , Cohen's  $d = 2.27$ , 36 ms,  $t(50) = 3.64, p = 0.001$ , Cohen's  $d = 1.06$ , and 48 ms,  $t(50) = 3.1, p = 0.003$ , Cohen's  $d = 0.9$ ; no SOAs  $> 48$  ms (including the no-mask condition) showed significant differences between the two mask  $\alpha$  groups ( $ps \geq 0.164$ ). Thus, the effect of noise mask  $\alpha$  (i.e., stronger masking for amplitude spectra  $\alpha = 1.0$ ) was only found within the critically important first  $\sim 50$  ms of scene gist processing time (Loschky et al., 2007; Loschky et al., 2010), which followed the target-mask amplitude slope similarity principle observed in Experiment 1.

## Experiment 3

So far, we have found that amplitude-only slope  $\alpha = 1.0$  masks produce the strongest backward masking of

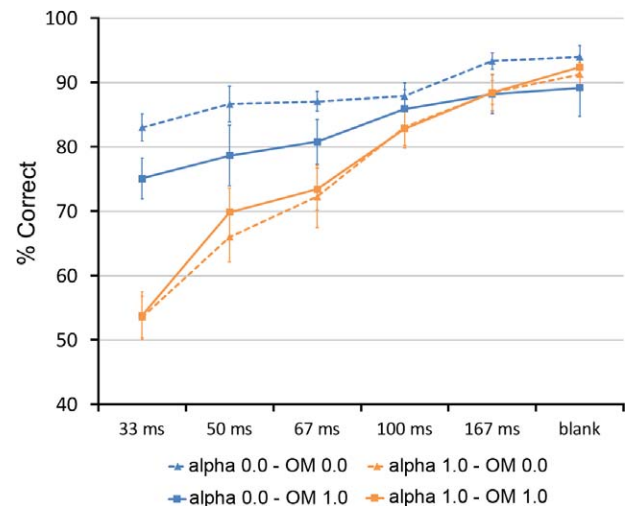


Figure 3. Averaged data from Experiment 2. On the ordinate is averaged percent correct (note that the scale begins at 40%), and on the abscissa is SOA. Error bars are  $\pm 1$  SEM.

rapid scene categorization (when compared to  $\alpha = 0$  [white noise], 0.5, and 1.5), specifically very early in scene category processing (i.e., the first  $\sim 50$  ms of processing time). One simple explanation for the stronger masking produced by  $\alpha = 1.0$  masks is that they have higher perceived contrast than smaller or larger  $\alpha$  values and higher perceived contrast produces stronger masking. Specifically, previous studies (e.g., Cass, Alais, Spehar, & Bex, 2009; Field & Brady, 1997; Tolhurst & Tadmor, 2000) have noted a difference in subjective contrast between rms contrast balanced images with varying  $\alpha$ . More specifically, Hansen and Hess (2012; figure 4a) found that  $\alpha = 1.0$  noise patterns were perceived to have 83% more rms contrast than  $\alpha = 0.0$  noise and  $\sim 50\%$  more rms contrast than  $\alpha = 1.5$  noise patterns. This trend mirrors the results of Experiment 1, in which the largest advantage for  $\alpha = 1.0$  noise masks was in comparison to  $\alpha = 0.0$  noise masks, and the next largest advantage was in comparison to  $\alpha = 1.5$  noise masks.

Thus, Experiment 3 sought to equate the perceived contrast of the  $\alpha = 0.0$  and  $\alpha = 1.0$  noise masks. The contrast matching data reported by Hansen and Hess (2012) found a  $\sim 83\%$  difference in perceived contrast between  $\alpha = 0.0$  and  $\alpha = 1.0$  noise. Thus, in Experiment 3, we set the  $\alpha = 0.0$  [white] noise masks to an rms contrast twice the rms contrast of the  $\alpha = 1.0$  masks.

## Method

### Participants

Thirty-two Kansas State University students (16 female), all naïve to the purpose of the study, participated for course credit after giving their Institutional Review Board-approved informed written

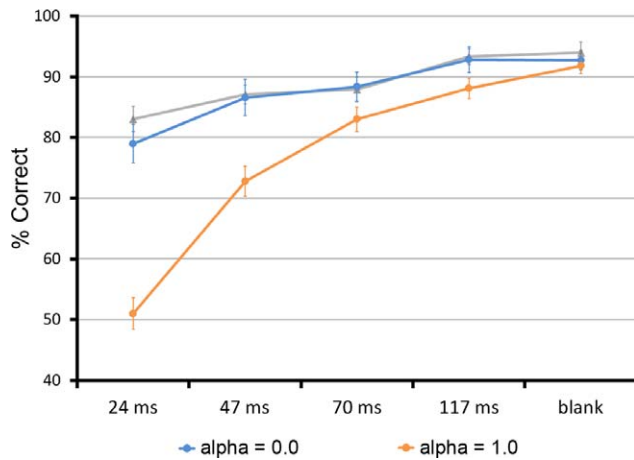


Figure 4. Averaged data from Experiment 3. On the ordinate is averaged percent correct (note that the scale begins at 40%), and on the abscissa is SOA. The gray trace plots data from the noise mask  $\alpha = 0.0$  and  $O_{\text{mag}} = 0.0$  from Experiment 2 ( $rms = 0.126$ ). Error bars are  $\pm 1$  SEM.

consent (with written parental consent also given for the subject under the age of 18). All observers had normal (or corrected to normal) vision and their ages ranged from 17 to 23 years ( $M = 19.1$ ,  $SD = 1.42$ ).

### Design and procedure

The mask stimuli were identical to those used in Experiment 2, but with the following exceptions. There were only two mask amplitude spectrum  $\alpha$  values ( $\alpha = 0.0$ ,  $rms = 0.252$ , and  $\alpha = 1.0$ ,  $rms = 0.126$ ) and one mask orientation bias magnitude (0.0, no orientation bias). Target stimuli were drawn from the set described in the General method section.

The current experiment employed the same paradigm as Experiment 2. The experiment used a 2 (Mask  $\alpha$ )  $\times$  5 (Target/Mask SOA) mixed design, with mask  $\alpha$  being the between-subjects variable and SOA being the within-subjects variable. Within a given mask  $\alpha$  and orientation bias magnitude condition, each participant saw 60 different target category stimuli for each SOA (10 stimuli randomly selected without replacement from each scene category). For the no-mask condition, an additional 60 images were randomly sampled (10 per scene category), for a total of 360 trials. Order of target category and SOA (including no mask) was randomized for each trial. Participants were randomly assigned to the between-subjects condition. The trial sequence was identical to that reported in Experiment 2 (including practice trials).

### Results and discussion

Random assignment of participants to the between-subjects variable, mask amplitude spectrum slope  $\alpha$ ,

resulted in 13 in the  $\alpha = 0.0$  condition and 19 in the  $\alpha = 1.0$  condition. Figure 4 shows mean scene categorization accuracy for each amplitude spectrum slope,  $\alpha$ , condition as a function of target/mask SOA (and no mask), and we conducted a 2 (Mask  $\alpha$ )  $\times$  5 (Target/Mask SOA) mixed repeated measures ANOVA to test the effects of these factors. As can be seen in Figure 4, there was a large difference between the two mask  $\alpha$  conditions, which was confirmed a significant between-subjects main effect of mask  $\alpha$ ,  $F(1, 30) = 14.47$ ,  $p < 0.001$ , Cohen's  $f^2 = 0.21$ . This is despite the fact that the  $\alpha = 0.0$  masks possessed twice as much physical contrast as the  $\alpha = 1.0$  masks and were approximately equivalent in perceived contrast.

We next carried out post-hoc independent  $t$  tests between the two mask  $\alpha$  groups at each SOA. For the five comparisons, the Bonferroni adjusted family-wise  $\alpha$  level was set to 0.01. In order to carry out the independent groups two-tailed  $t$  tests, we first created equal sized groups by randomly selecting six subjects to omit from the  $\alpha = 1$  condition (though the statistical test results were the same whether the subjects were omitted or not). As shown in Figure 4, the comparisons with SOAs of 50 ms or less were statistically significant ( $t_s \geq 3.95$ ,  $p_s \leq 0.001$ , Cohen's  $d_s \geq 0.83$ ), but not from 70 ms SOA onward, SOA = 70 ms:  $t(24) = 2.21$ ,  $p = 0.037$ ; SOA = 117 ms:  $t(24) = 2.15$ ,  $p = 0.042$ ; no mask:  $t(24) = 0.77$ ,  $p = 0.450$  (based on the family-wise adjusted  $\alpha$  level).

Thus, Experiment 3 showed that the difference in masking between the  $\alpha = 1.0$  and 0.0 conditions in Experiments 1 and 2 could not be eliminated by simply matching the perceived contrast of the masks (by doubling the physical contrast of the  $\alpha = 0.0$  masks). Even more strikingly, as shown in Figure 4, the  $\alpha = 0.0$  masking function with  $rms$  contrast = 0.252 did not significantly differ from that in Experiment 2 with  $rms$  contrast = 0.126 ( $p > 0.05$ ). Again, masking was greatest for target/mask SOAs of roughly 50 ms or less as in Experiment 2, and continued to follow the target-mask amplitude slope similarity principle observed in Experiment 1.

## Experiment 4

Experiments 1–3 established that amplitude-only mask slope  $\alpha$  was the crucial determining factor modulating masking of rapid scene categorization. Interestingly, that modulation followed a target-mask amplitude slope similarity principle in which the more similar the mask image slope was to the target image slope, the stronger the masking. Here, we set the stage to examine whether and how unrecognizable amplitude + phase-defined mask structure masks rapid scene categorization differently from amplitude-only defined structure. Specifically, does it modulate masking effects in a category-specific manner? The current experiment

provided key information about the nature of the masks used in Experiment 5 to answer this question.

Given these aims, it is important to hold the amplitude-defined mask structure approximately constant while varying amplitude + phase defined structure (mainly phase-defined structure, i.e., present vs. absent). Experiments 1–3 all showed amplitude spectrum slope producing very strong masking modulation (ranging from  $\sim 55\%$  accuracy to  $\sim 85\%$  accuracy), so we chose to test for category-specific masking effects with a slope  $\approx 1.0$  (typical of natural scenes). Thus, all masks were created to fall within a 0.12 standard deviation around a 1.12 slope, namely equal to the mean and standard deviation of slope  $\alpha$  of our target image set.

Since we were interested in assessing the role of unrecognizable amplitude + phase-defined structure in masking rapid scene categorization, we carried out the current experiment to determine the recognizability of the images that we planned to use as masks in Experiment 5 (to select masks with the lowest recognizability). This was the same rationale used by Loschky et al. (2007) and Loschky et al. (2010) for measuring the recognizability of mask images. Specifically, if one type of masking image is more recognizable than another, and that mask type is also shown to cause greater masking, then the difference in masking could be attributed to conceptual masking (Intraub, 1984; Loftus & Ginn, 1984; Loftus, Hanna, & Lester, 1988; Potter, 1976), rather than the masks' image statistical properties. Thus, by choosing masks with low recognition probability, we can strongly minimize any influence of conceptual masking.

## Method

### Participants

Fifty-six Kansas State University students (27 female), all naïve to the purpose of the study, participated for course credit after giving their Institutional Review Board-approved informed written consent. All observers had normal (or corrected to normal) vision and their ages ranged from 18 to 30 years ( $M = 19.60$ ,  $SD = 2.19$ ).

### Mask construction

Each of the six image category sets consisted of 100 images. Of those, 24 images were randomly selected to be used as target images (i.e., for use in Experiment 5), and the remaining 76 images within each set were used to generate the masking stimuli for that basic level category (i.e., for use in Experiment 4 and in Experiment 5). Experiment 4 tested the recognizability of two types of visual mask stimuli, called AMP masks (i.e., amplitude-only masks) and AMP + PHASE

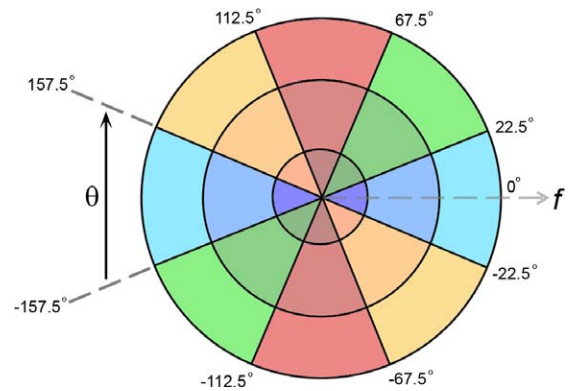


Figure 5. Illustration of sectors and sector segment divisions in Fourier space (referenced in polar coordinates). The dashed gray arrow shows the spatial frequency,  $f$ , dimension, and the solid black arrow shows the orientation,  $\theta$ , dimension. Color represents a full sector “bow tie” for each orientation, with color saturation representing spatial frequency segments. Note that spatial frequency segments are scaled for ease of visibility and do not represent actual defined spatial frequency area in Fourier space.

masks, to be used in Experiment 5. Both mask types were created to possess identical amplitude spectral relationships, but differ in the AMP + PHASE masks possessing scene image features largely defined by the phase spectrum (e.g., scene-specific edges and lines). Below we describe how each mask type was constructed, based on methodology modified from Hansen and Hess (2007), as illustrated in Figures 5–7.

The AMP + PHASE masks were created on a category-by-category basis, so that, for example, the airport category masks only included phase information from airport images (specifically, from the 76 images not selected as target stimuli). AMP + PHASE masks were created by selectively sampling defined portions of the amplitude and phase spectra of 12 randomly selected seed images from the remaining 76 images within the given category set. Thus each mask was derived from 12 different seed images within a category set. Once a given mask was generated, its corresponding seed images were returned to the image pool for the given category set, and thus any image could be resampled to create another mask stimulus. Thus, the seed images were never repeated within a given mask stimulus, but could be randomly resampled for a different mask stimulus from the same category set.

Each AMP + PHASE mask was constructed by starting with two empty composite spectra (i.e., both filled with zeros) in the Fourier domain, consisting of two  $512 \times 512$  matrices. Each matrix was referenced in polar coordinates, and split into sector segments, illustrated in Figure 5. Specifically, each sector was defined by a  $45^\circ$  band of orientations,  $\theta$ , centered on

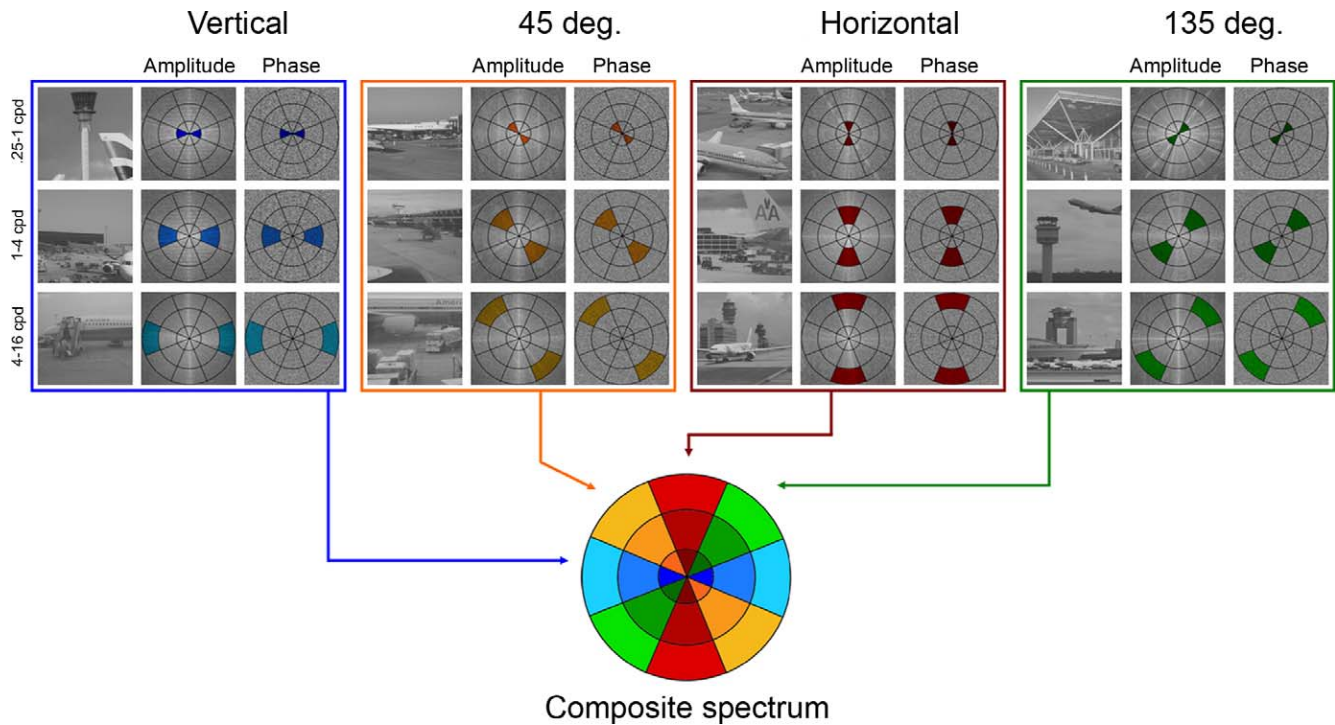


Figure 6. Illustration of creating a given composite spectrum using scenes from the airport category in Experiments 4 and 5. The blue bow tie in Fourier space represents vertical image structure in image space. Three separate airport scene images (shown in the far left of the blue box) are used to construct that bow tie. To the right of those images (middle column) are the corresponding amplitude spectra and to the right of those, the corresponding phase spectra. The lower spatial frequencies (0.25–1.0  $c/^\circ$ ) are taken from the top image, the intermediate frequencies (1.0–4.0  $c/^\circ$ ) from the middle image, and the high frequencies from the bottom image. The process is repeated for the three remaining bow ties, but with different images for each bow tie. In total, 12 randomly selected airport images would be used to create a composite AMP + PHASE airport image.

four primary orientations, namely vertical ( $0^\circ$ ),  $45^\circ$  oblique, horizontal ( $90^\circ$ ), and  $135^\circ$  oblique. The sectors were mirrored to the polar opposite side of the Fourier domain to form a “bow tie” reference system for each primary orientation (as illustrated with the color bow ties in Figures 5 and 6). Sector segments were defined as three 2-octave bands of spatial frequencies,  $f$ , within each sector (and mirrored to the polar opposite side of the Fourier domain). Sector segments were defined according to  $f$  and extended from 0.25–1  $c/^\circ$ , 1–4  $c/^\circ$ , and 4–16  $c/^\circ$ . This reference system yields 12 mirrored sector segments, one for each of the four orientation bands and each of the three spatial frequency bands.

To construct an AMP + PHASE mask, we filled its empty composite spectrum. First, we randomly selected one of the 12 seed images in a scene category and ran a DFT of it, which yielded its amplitude spectrum,  $A(f_i, \theta_j)$ , and phase spectrum,  $\Phi(f_i, \theta_j)$ . Then, we randomly selected a mirrored sector segment from  $A(f_i, \theta_j)$  and  $\Phi(f_i, \theta_j)$ , with the mirrored sector segments taken from  $A(f_i, \theta_j)$  and  $\Phi(f_i, \theta_j)$  being identical in terms of location. The amplitude coefficients and phase angles within the selected mirrored segment of  $A(f_i, \theta_j)$  and  $\Phi(f_i, \theta_j)$  were copied into their corresponding mirrored sector segments of the composite amplitude and phase

spectrum, as illustrated in Figure 6. For example, in Figure 6, the top-left corner represents the mirrored sector segment containing spatial frequencies between 0.25–1  $c/^\circ$  centered on vertical. To get this information, we randomly selected a seed image, and assigned its corresponding amplitude coefficients from  $A(f_i, \theta_j)$  and

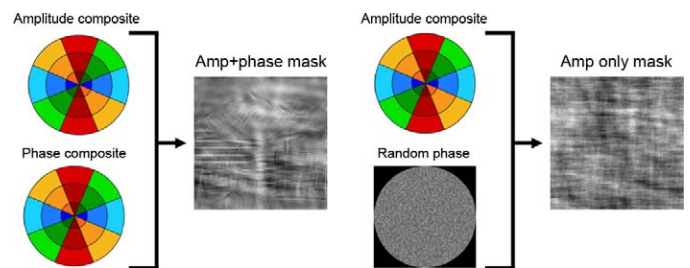


Figure 7. Illustration showing the contribution of the different composite spectra for an amp + phase mask (left) and an amp mask (right) in Experiments 4 and 5. Thus, an amp + phase mask is made up of composite amplitude and phase spectra sampled from 12 different images (see Figure 6), while the corresponding amp mask uses only the composite amplitude spectrum combined with a randomized phase spectrum. Thus, the two masks shown above only differ with respect to their phase spectra (i.e., their amplitude spectra are identical).

phase angles from  $\Phi(f_i, \theta_j)$  to that location in the composite phase and amplitude spectrum. We repeated this process until all of the 12 mirrored sector segments of the composite phase and amplitude spectrum were filled. We then squared the composite amplitude spectrum (to get the power), and normalized it to the average power spectrum of the entire normalized scene category images. We assigned the DC (i.e., the zero frequency component) of the same averaged power spectrum to the squared composite amplitude spectrum. These last two steps ensured that each AMP + PHASE mask had the identical mean luminance and rms contrast of the target images. Finally, we rendered the AMP + PHASE masks in the spatial domain by taking the discrete inverse fast Fourier transform of composite amplitude and phase spectra, with a resulting AMP + PHASE image shown in Figure 7 (left half). Refer to Figures 6 and 7 for further details.

Figure 7 (right half) shows how we generated the AMP masks during the process of creating the AMP + PHASE masks. Specifically, for any given AMP mask, instead of taking the DFT of the composite amplitude and phase spectra, we replaced the composite phase spectrum with a randomized phase spectrum. Furthermore, we used the same randomized phase spectrum for each mask stimulus across all scene image categories. We constructed a given random phase spectrum by assigning random values from  $-\pi$  to  $\pi$  to the coordinates of a  $512 \times 512$  polar matrix,  $\Phi_{RAND}(f, \theta)$ , such that the phase spectra were odd-symmetric—i.e., for  $\theta$  angles in the  $[\pi, 2\pi]$  half of polar space,  $\Phi_{RAND}(f, \theta) = -\Phi_{RAND}(f, \theta)$ . Refer to Figure 7 for further details. Figure 8 shows example masks from each basic level scene category.

As described above and illustrated in Figures 6 and 7, the AMP + PHASE masks differed across categories by both amplitude and phase, whereas the AMP masks differed only with respect to amplitude. We also note that the technique we described above for creating both mask types essentially constitutes a rudimentary wavelet methodology. Specifically, as in typical wavelet methods, composite spectra were constructed from relatively narrow bands of spatial frequency and orientation. However, unlike typical wavelet methods, these composite global spectra were built up from local portions of different images' Fourier spectra. This wavelet-like approach differs from the standard approach to building amplitude-only masks or amplitude + phase masks, which has typically relied on applying global manipulations to either the amplitude spectrum (as done in Experiment 1), or systematically randomizing the phase spectrum (e.g., Loschky et al., 2007; Wichmann et al., 2006). The wavelet-like approach we used to create our AMP + PHASE masks means that they contained scene-specific features, such as lines and edges. However, the fact that we constructed the masks

from randomly mixed sectors of amplitude and phase spectra from different scenes means that the masks are likely unrecognizable (see Hansen & Hess, 2007, for further detail).

### Scene categorization task

The design of the current experiment consisted of a single-interval six-alternative forced-choice paradigm with two within-subject factors: 2 (Amplitude vs. Amplitude + Phase Masks)  $\times$  6 (Scene Categories of Masks). There were 28 mask images per cell, and each image was presented once, for a total of 336 trials (randomly ordered). On each trial, participants saw a briefly flashed mask stimulus and indicated which of the six basic level scene categories they believed it belonged to. Each trial began with a fixation point (subtending  $0.31^\circ$  of visual angle) until the participant pressed a mouse button to initiate the trial. This was followed by a variable duration (300–900 ms) blank screen (neutral gray: luminance = mean of the entire target and mask image set), then a briefly flashed stimulus (either AMP + PHASE or AMP mask) for 24 ms (two refresh cycles at 85 Hz), then a blank screen (same neutral gray) for 750 ms duration, then a response screen containing a  $2 \times 3$  grid of the six basic level scene category labels, until the participant responded by a mouse click. As in Experiments 1–3, the position of each of the category labels in the response label grid was randomized on every trial.

### Results and discussion

One participant's data was removed from the analyses for only responding "Forest" on every trial (except one, in which they correctly responded "Street").

Results for mean accuracy by image type and basic level category are shown in Figure 9a. As shown, accuracy for four of the six categories was at or slightly below chance (16.67%). However, two of the categories, beach and forest, were more accurately identified than the others. This was verified by a 2 (Mask Type: AMP vs. AMP + PHASE)  $\times$  6 (Scene Category) factorial repeated measures ANOVA on accuracy, which showed a main effect of category,  $F(5, 265) = 21.812$ ,  $p < 0.001$ , Cohen's  $f^2 = 0.401$ . Follow-up Bonferroni-corrected  $t$  tests (adjusted alpha = 0.0083) showed that beach,  $t(54) = 5.029$ ,  $p < 0.001$ , Cohen's  $d = 0.684$ , and forest,  $t(54) = 6.182$ ,  $p < 0.001$ , Cohen's  $d = 0.841$ , were significantly above chance, and that store,  $t(54) = -5.843$ ,  $p < 0.001$ , Cohen's  $d = 0.795$ , was significantly below chance. The other categories did not reach significance.

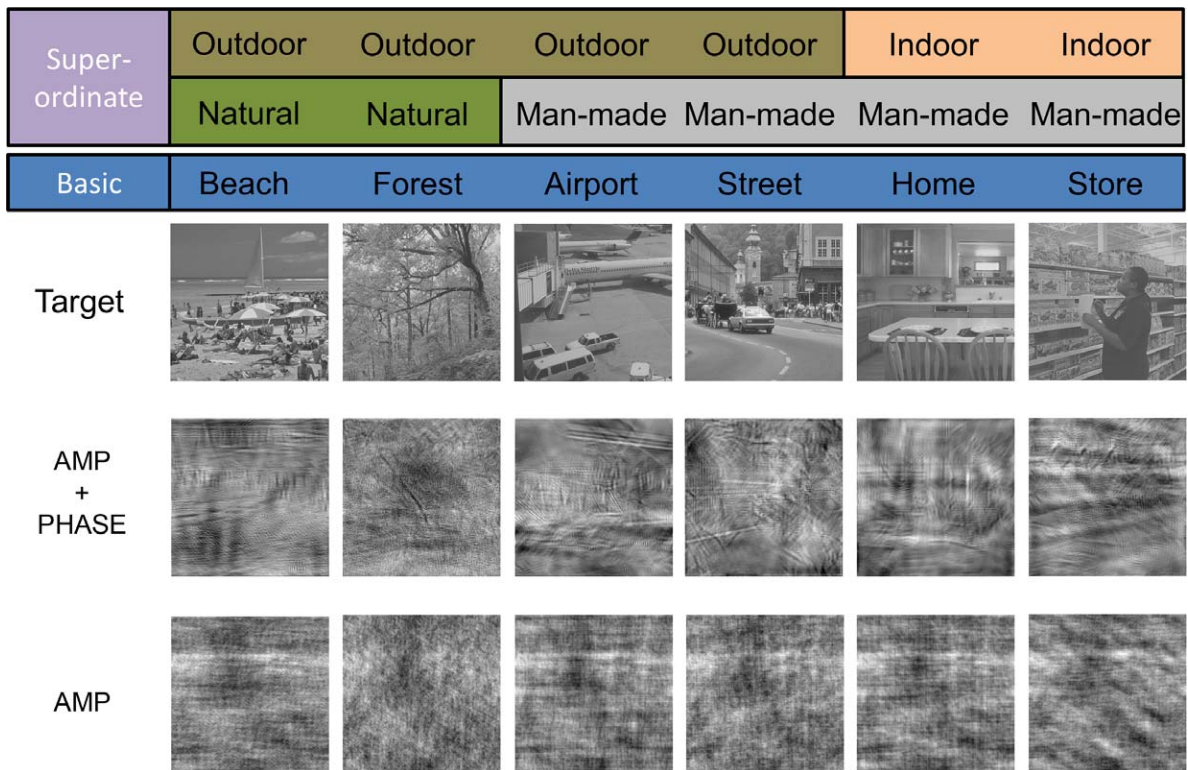


Figure 8. Example target and mask stimuli used in Experiment 4 (mask images only) and Experiment 5 (target and masks). The top three text rows show the categorization hierarchy, with the top two text rows showing the superordinate category structure and the bottom text row showing the basic-level category structure. The three rows of images show example targets and masks (AMP + PHASE and AMP) from each of the basic-level categories.

Overall, these results can largely be explained in terms of a response bias to more frequently select natural categories (a natural bias) (Joubert et al., 2009; Loschky & Larson, 2008). To correct for this response bias, we calculated the percentage of correct category

responses out of each category’s response rate (Figure 9b). Stated more formally, we calculated  $p(\text{correct response} | \text{response} = X_{\text{cat}}) / p(\text{response} = X_{\text{cat}})$ , where  $X_{\text{cat}}$  = one of the six categories. Bonferroni-corrected  $t$  tests (adjusted alpha = 0.0083) showed only three instances

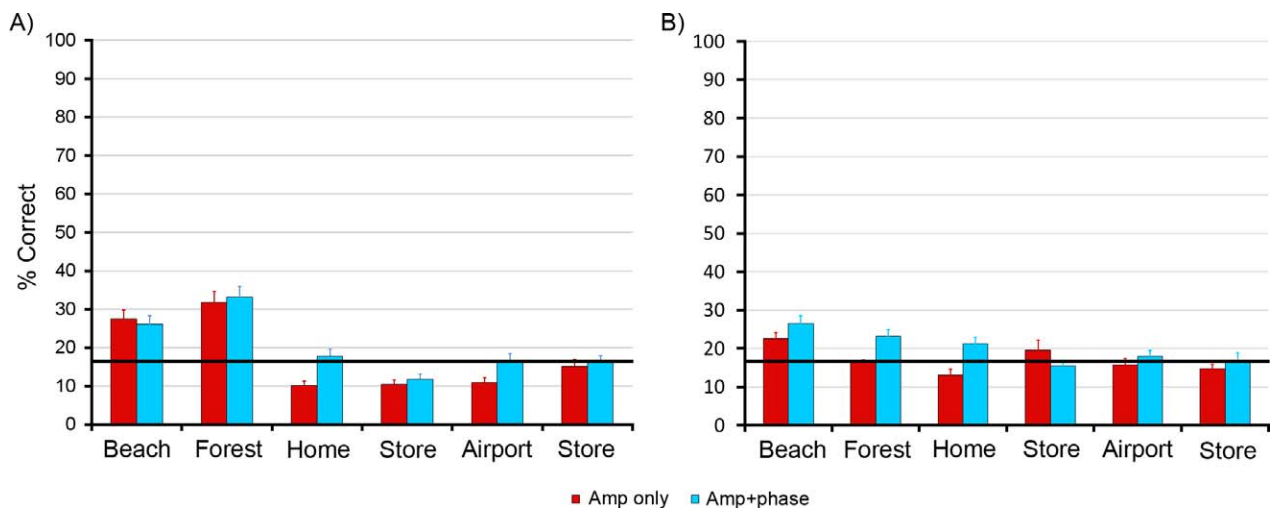


Figure 9. (A) Averaged data from Experiment 4 showing recognition accuracy (ordinate) for amplitude and phase masks as a function of scene category (abscissa). (B) Data from Experiment 4 showing percentage of correct category responses out of total category response rate. The black line marks chance performance in both panels, and all error bars are  $\pm 1$  SEM.

of above-chance correct category response percentages: AMP masks beach,  $t(53) = 3.51$ ,  $p < 0.001$ , Cohen's  $d = 0.474$ , and AMP + PHASE masks beach,  $t(53) = 4.69$ ,  $p < 0.001$ , Cohen's  $d = 0.633$ , and forest,  $t(53) = 3.70$ ,  $p < 0.001$ , Cohen's  $d = 0.499$ . We will return to this result in Experiment 5.

## Experiment 5

Experiments 1–3 showed that amplitude-only masks produced performance largely dependent on the amplitude slope, following a target-mask amplitude spectrum slope similarity principle. In Experiment 5, we investigated whether amplitude + phase masks would yield masking effects that differed from amplitude-only masks, and if so, how? The results of Experiment 1 (and Loschky et al., 2007) suggested that amplitude-only masks do not yield category-specific masking effects. Here, we hypothesized that if category-specific features are used to process scenes to categorization, then target-mask categorical similarity should affect scene gist masking—namely, whether the target and mask categories are the same or different at the superordinate level, or at the basic level, should affect the degree masking of rapid scene categorization. Furthermore, since the masks generated in Experiment 4 possessed amplitude slopes,  $\alpha$ , that all fell within 0.12 standard deviation of 1.12 (i.e., essentially  $\alpha = 1$ ), if amplitude + phase masks did not mask differently than observed in Experiments 1–3, than we would not expect to observe any category-specific masking effects.

## Method

### Participants

Thirty Kansas State University students (14 female), all naïve to the purpose of the study, participated for course credit after giving their Institutional Review Board-approved informed written consent (with written parental consent also given for those under the age of 18). All observers had normal (or corrected to normal) vision and their ages ranged from 17 to 45 years ( $M = 20.5$ ,  $SD = 4.83$ ).

### Stimuli

The stimuli were those used in Experiment 4. There were 144 target images (24 Images  $\times$  6 Scene Categories). There were also 144 mask images, 72 of each type (AMP and AMP + PHASE masks), and 12 in each of six scene categories.

## Design and procedure

Experiment 5 involved four factors: 2 (AMP vs. AMP + PHASE masks)  $\times$  6 (Scene Categories of Targets)  $\times$  6 (Scene Categories of Masks)  $\times$  2 (Valid/Invalid Category Labels) = 144 cells in the design. There were two trials per cell, for a total of 288 trials (randomly ordered). There were 24 images per scene or mask category, for a total of 144 target images and 144 mask images. Each target was used twice in the experiment, once validly labeled, and once invalidly (falsely) labeled. Each mask was presented twice. Each image was also masked once with an AMP mask and once with an AMP + PHASE mask.

The general procedure was identical to that in Experiments 1–3, with the following exceptions. First, the current experiment used only a single masking SOA (36 ms) based on a target duration of 24 ms, and an ISI of 12 ms. The mask was 72 ms duration, for a 3:1 mask:target duration ratio. Second, the response options at the end of a trial differed. Following the presentation of the target, mask, and blank screen, a single category label was presented until the participant made either a “Yes” or “No” response. The category labels were valid (matched the presented image) 50% of the time and invalid 50% of the time. There were 288 trials, with each of the six category labels presented equally often.

## Results and discussion

Nine trials across 30 subjects were found to have longer target, ISI, or mask durations than intended, and were removed from the analyses, leaving a total of 8,631 trials.

We first analyzed our data in cases in which the target and mask came from different categories. Of particular interest was whether the AMP versus AMP + PHASE masks caused differential rapid scene categorization masking, since these masking conditions differed in terms of having phase-defined image features, with AMP masks possessing no phase-defined structure (compare the bottom two rows of Figure 8). To answer this question, our first analysis was a 2 (Mask Type: AMP vs. AMP + PHASE)  $\times$  6 (Category) repeated-measures ANOVA on rapid scene categorization accuracy to determine any general differences in masking strength. The results are shown in Figure 10a.

As shown in Figure 10a, the AMP + PHASE masks interfered more with rapid scene categorization accuracy than the AMP masks,  $F(1, 29) = 15.062$ ,  $p = 0.001$ , Cohen's  $f^2 = 0.122$ . This suggests that unrecognizable amplitude + phase-defined image characteristics disrupt rapid scene categorization more than amplitude-defined characteristics alone. As also shown in Figure 10a, there was a main effect of mask category,  $F(5, 145)$

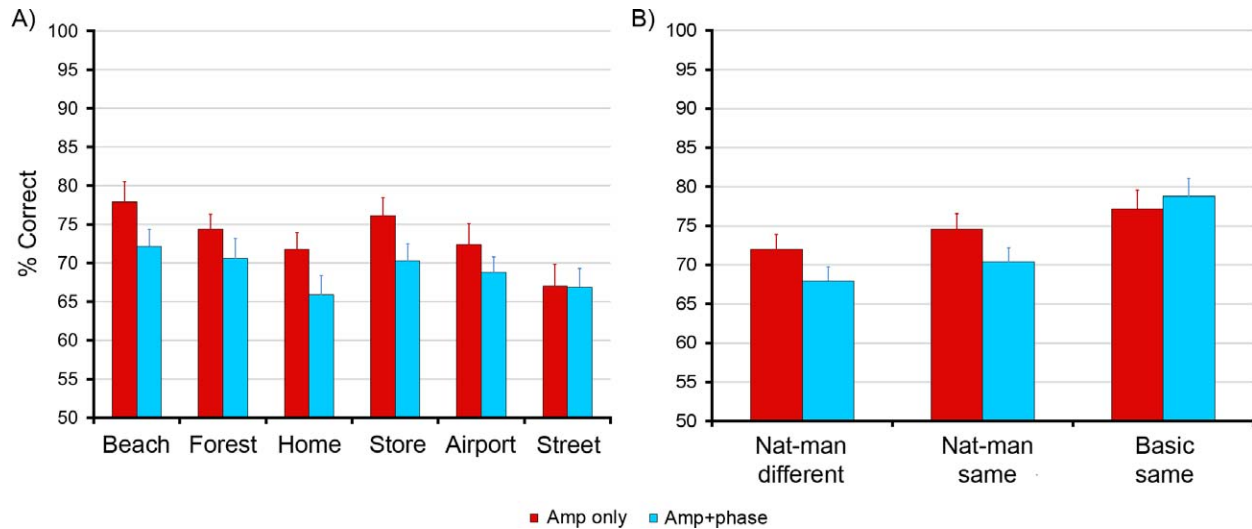


Figure 10. (A) Averaged data from Experiment 5 showing rapid scene categorization accuracy (ordinate) as a function of mask type (amplitude vs. phase masks) and mask category (abscissa) (note that the ordinate scale begins at 50% [= chance] to increase visibility). (B) Averaged data from Experiment 5 data replotted to show rapid scene categorization accuracy (ordinate) as a function of mask type (amp only vs. amp + phase masks) and target-mask categorical similarity (abscissa) (note that the ordinate scale begins at 50% [= chance] to increase visibility). Error bars are  $\pm 1$  SEM. “Nat-man different” = target and mask are different in terms of the natural/man-made distinction; “nat-man same” = target and mask are the same in terms of the natural/man-made distinction, but different at the basic level; “basic same” = target and mask are the same in terms of the basic level. Error bars are  $\pm 1$  SEM.

$= 4.818$ ,  $p < 0.001$ , Cohen’s  $f^2 = 0.193$ , with some mask categories, such as beach and forest masks, causing less masking than others, such as street masks. Note that this runs exactly opposite to the conceptual masking hypothesis, since the most recognizable mask categories, beach and forest (as shown in Experiment 4), caused among the least disruption of rapid scene categorization when used as masks. In addition, mask type did not significantly interact with mask category,  $F(5, 145) < 1$ ,  $p = 0.456$ . The above results therefore support the notion that category-specific phase-defined mask structure produces differential scene gist masking (here, in the form of masking strength) than amplitude-only defined mask structure. However, the question still remains as to whether such signals operated in a category-specific manner. Thus, the other aim of the Experiment 5 was to determine whether target-mask category similarity would affect rapid scene categorization—namely, do amplitude + phase-defined category-specific features selectively mask rapid scene categorization?

To address the above question, we analyzed our data in terms of both the superordinate and basic level categorical similarity of the target and mask images, using the following similarity scale: natural/man-made different (e.g., beach target & street mask) < natural/man-made same, basic different (e.g., beach target & forest mask) < basic level same (e.g., beach target & beach mask). We carried out a 2 (Mask Type: AMP vs. AMP + PHASE)  $\times$  3 (Categorical Similarity: Natural/Man-Made Different vs. Natural/Man-Made Same,

Basic Different vs. Basic Level Same) repeated measures ANOVA on scene categorization accuracy.

As shown in Figure 10b, the more categorical dissimilarity between target and mask images the greater the interference with rapid scene categorization accuracy,  $F(2, 58) = 27.36$ ,  $p < 0.001$ , Cohen’s  $f^2 = 0.541$ . Specifically, Figure 10b shows that the least categorically similar target-mask pairings (nat-man different) showed the strongest masking (lowest accuracy). Planned comparisons showed that whether target and mask were the same or different at the superordinate level (nat-man different vs. nat-man same, basic different) significantly affected categorization accuracy,  $F(1, 29) = 5.367$ ,  $p = 0.028$ , Cohen’s  $f^2 = 0.270$ , but that the biggest difference in masking strength was due to whether target and mask were the same or different at the basic level (nat-man same, basic different vs. basic same),  $F(1, 29) = 22.422$ ,  $p < 0.001$ , Cohen’s  $f^2 = 0.598$ . As before, AMP + PHASE masks caused significantly greater masking than AMP masks,  $F(1, 29) = 4.377$ ,  $p = 0.045$ , Cohen’s  $f^2 = 0.137$ . Interestingly, it appears in Figure 10b that categorical dissimilarity between target and mask images mattered more for the AMP + PHASE masks, and less so for the AMP masks; however this crossover interaction just failed to reach significance when correcting for non-homogeneity of variance,  $F(1.664, 48.252$  [Huynh-Feldt]) = 3.356,  $p = 0.052$ , Cohen’s  $f^2 = 0.162$ . Because this was so close to being significant, we carried out two one-way ANOVAs for categorical similarity, for AMP + PHASE masks and for AMP masks. In fact, as shown in Figure 10b,



there was a significant main effect for categorical dissimilarity for both types of masks; however, the effect size was approximately twice as large for the AMP + PHASE masks,  $F(2, 58) = 19.404$ ,  $p < 0.001$ , Cohen's  $f^2 = 0.640$ , as for the AMP masks,  $F(2, 58) = 5.84$ ,  $p = 0.005$ , Cohen's  $f^2 = 0.328$ . It therefore appears that both types of mask follow a *target-mask categorical dissimilarity principle* (i.e., the more dissimilar the target-mask pairs were in terms of category, the stronger the masking), with AMP + PHASE masks yielding the strongest effects.

Importantly, if the effects shown in Figure 10b were due to the small level of mask recognizability (Experiment 4), they are very inconsistent with the predictions of conceptual masking. Specifically, the conceptual masking hypothesis predicts greater masking by more recognizable mask categories (Intraub, 1984; Loftus & Ginn, 1984; Loftus, Hanna, & Lester, 1988; Michaels & Turvey, 1979; Potter, 1976), whereas we found that the more recognizable categories caused less masking. Thus, the categorical dissimilarity effect observed here cannot be accounted for by conceptual masking. Curiously, the AMP masks here in Experiment 5 produced a modest categorical dissimilarity effect, while the amplitude-only masks in Experiments 1–3 yielded no category-specific effects. One possible explanation may have to do with the methodology employed to construct the different types of amplitude-only masks (i.e., one consisting of global transformations, while the other resulted from rudimentary wavelet techniques). We will return to this issue in the General discussion.

## General discussion

This study investigated the ability of amplitude-only-defined and amplitude + phase-defined unrecognizable image structure to mask rapid scene categorization performance, in particular (a) whether the two types of masks would yield different masking functions, and if so (b) how the different mask spectral characteristics would modulate the masking functions. Experiments 1–3 showed a greater impact on scene categorization of amplitude spectrum slope,  $\alpha$ , than orientation at processing times  $< 50$  ms, which followed a target-mask amplitude slope similarity principle (i.e., stronger masking when target and mask amplitude spectrum slopes were more similar). Experiment 5 then showed a greater impact on scene categorization of unrecognizable amplitude + phase-defined image structure than amplitude-only-defined structure. Further, both types of masking effects in Experiment 5 followed a target-mask categorical dissimilarity principle (i.e., stronger masking effects when target and mask scene categories

differed more), but with amplitude + phase masks showing effect sizes that were almost twice those of the amplitude-only masks. Interestingly, the amplitude-only masks in Experiments 1–3 produced no category-specific masking effects, whereas Experiment 5's amplitude-only masks produced modest category-specific masking effects. One possible reason for this difference may have been the differences in mask construction between Experiments 1–3 versus 4–5. Below, we discuss the results of Experiments 1–3 and Experiments 4–5 in greater detail, in an effort to provide a qualitative account of the possible underlying mechanisms regulating both types of masking effects.

The results from Experiments 1–3 showed that the amplitude slope is the dominant factor in producing masking effects by phase-randomized scenes (with  $\alpha = 1.0$  producing the largest masking effect). Further, the interference caused by  $\alpha = 1.0$  masks relative to  $\alpha = 0.0$  masks (regardless of orientation bias) occurs within the first 50 ms of processing (and cannot be explained by differences in perceived contrast). Such an early modulation suggests that the relevant interactions took place very early in the cortical processing stream, possibly within striate cortex (i.e., V1). Given the possibility of an early neural substrate for the effects observed in Experiments 1–3, and the large differences in contrast at different spatial frequencies as  $\alpha$  was varied, it would be tempting to explain the masking effects by differences in spatial frequency-specific contrast differences (i.e., a simple spatial channel account). Specifically, while the noise masks and scene target images were equated for rms contrast, masks with  $\alpha < 1$  or  $\alpha > 1$  would have spatial frequency-specific contrast differences from the target image set (having averaged  $\alpha = 1.12$ ,  $SD = 0.12$ ). Thus, noise masks with  $\alpha = 0.0$  contained more luminance contrast at higher spatial frequencies than the target image set, while noise masks with  $\alpha = 1.5$  had more luminance contrast at lower spatial frequencies (see Figure 1b). However, neither  $\alpha = 0.0$  or  $1.5$  produced the largest masking effects, so the effects reported in Experiment 1 cannot be explained by contrast differences at either the lowest or highest spatial frequencies. Further, the  $\alpha = 0.0$  noise masks in Experiment 3 had an rms contrast twice that of the  $\alpha = 1.0$  noise masks, yet the masking strength advantage for the  $\alpha = 1.0$  noise masks within the first 50 ms persisted.

Since the masking differences in Experiments 1–3 imply an early neural substrate for those effects, yet cannot be explained by contrast biases at the lowest or highest spatial frequencies, the question as to why  $\alpha = 1.0$  noise masks produce the largest masking effects (i.e., the target-mask slope similarity effect) remains. Interestingly, the modulation of masking strength by mask  $\alpha$  in the current study is virtually identical to the simultaneous masking effects observed by Hansen and

Hess (2012) in which participants had to identify the orientation of Gabor patches or identify letters. There,  $\alpha = 1.0$  noise masks were found to produce stronger masking than  $\alpha = 0.0$  or  $1.5$  noise masks. Hansen and Hess (2012) argued that the different  $\alpha$  masking strengths could be explained by modeled cortical interactions employing contrast gain control processes (e.g., Carandini & Heeger, 1994; Heeger, 1992a, 1992b; Wilson & Humanski, 1993), which take place exclusively within V1. Their contrast gain control account built on the work of David Field and Nuala Brady (Brady & Field, 1995; Field, 1987; Field & Brady, 1997) who proposed a modified multichannel model of V1. That model predicts overall larger spatial frequency channel responses for  $\alpha = 1.0$  stimuli (compared to stimuli with smaller or larger  $\alpha$ s). In their model, the spatial frequency bandwidth of different early visual channels is held constant in octaves on a log axis, with the peak sensitivity of each filter channel also held constant, based on the results of a large sample of striate neurons (R. L. De Valois, Albrecht, & Thorell, 1982). With such a configuration, any image possessing an amplitude spectrum slope  $\alpha \approx 1.0$  will produce equivalently large contrast energy responses across all channels in an inhibitory contrast gain pool, regardless of the peak spatial frequency to which each filter is tuned (see Brady & Field, 1995, for further detail). Thus, if  $\alpha = 1.0$  noise largely and equally drives all spatial channels, a tuned gain pool for  $\alpha = 1.0$  noise should be much more active than with smaller or larger  $\alpha$ s, thus producing the greatest inhibitory effect on the output signal to perceptual decision processes. Furthermore, contrast gain control processes were shown to strongly operate during the first 50 ms of processing time in a backward masking paradigm (Essock, Haun, & Kim, 2009). Thus, the masking effects produced by the amplitude masks in Experiments 1–3 may simply reflect a variant of a contrast gain control modulated signal-to-noise ratio in early visual processes that depends much more on the distribution of contrast across spatial frequency than across orientation. Therefore, the target-mask slope similarity effect observed in Experiments 1–3 may result from a modified contrast gain control process implemented entirely in V1. The implication here is that the majority of amplitude-only masking effects may have more to do with specific early visual mechanisms than later scene categorization processes (e.g., in the PPA, the LOC, etc., Walther, Caddigan, Fei-Fei, & Beck, 2009). If so, such interactions would apply equally well to a wide array of visual tasks (e.g., Gabor orientation discrimination or letter recognition), rather than being limited only to rapid scene categorization. Additionally, the lack of any category-specific effects in Experiments 1–3 further supports the idea that those masking effects resulted from general neural operators, at least for

amplitude-only-defined content derived from global spectral manipulations.

Regarding the masking effects produced by unrecognizable amplitude + phase-defined image structure in Experiment 5, we observed that unrecognizable amplitude + phase masks interfered with rapid scene categorization in a manner much more specific to target-mask category pairs (i.e., Figure 10b), specifically following a target-mask categorical dissimilarity principle. In contrast to the account given for the mechanisms involved in the masking effects observed in Experiments 1–3, one cannot explain the target-mask categorical dissimilarity effect by a modified contrast gain control process. That is, because all masks in Experiments 4 and 5 were designed to contain approximately equivalent global contrast distributions across spatial frequency, the spatial channels involved in processing the masks would yield very similar output regardless of mask category. We verified this by passing all Experiment 5 masks through the bank of filters reported in Hansen and Hess (2012), and found that the greatest between-category difference for either mask type (AMP masks or AMP + PHASE masks) never exceeded a quarter of a log unit (across spatial frequency or orientation). Thus, if gain control modulated filter outputs did drive the masking effects reported in Figure 10b, the trend would be flat regardless of target-mask category pairing. Thus, neither spatial frequency- nor orientation-specific contrast change across the masks can account for the effects observed in Experiment 5. It is quite plausible that the greater accuracy when target and mask are from the same basic level category is due to integration of local category-specific image features from target and mask, which would be less disruptive (i.e., produce greater accuracy) when those features are more similar. Further research is needed to provide a definitive account.

Interestingly, Experiment 5 also yielded a modest target-mask dissimilarity effect when AMP masks were employed. This runs counter to the results of Experiments 1–3 (and Loschky et al., 2007, experiment 3). However, as noted in Experiment 4, the amplitude-only masks in Experiments 4 and 5 were constructed very differently from simple phase randomization. Specifically, they were constructed from rudimentary wavelet-like image sampling techniques (i.e., AMP masks were created by sampling different sector segments from a number of different image spectra) that may have resulted in spurious amplitude-only structure that signaled category-specific information. Further research is needed in order to address exactly how such an approach can lead to category specific masking. Nevertheless, such a difference in masking trends due to mask construction (global transformations vs. wavelet composition) therefore serves as a cautionary

note for future studies that seek to further examine the role of amplitude-only-defined structure in the masking of rapid scene categorization.

In sum, the current study's results strongly support the notion that rapid scene categorization masking effects differ depending on the masks' Fourier spectral properties. Further research is needed in order to determine if the slope similarity principle is tied directly to the slope of the target scene (here, our targets had an average slope of 1.12) or to the typical slope observed with natural scenes (i.e.,  $\alpha \approx 1.0$ ). Interestingly, Fourier amplitude spectrum slope produced the largest performance modulation with respect to categorization accuracy (ranging between  $\sim 55\%$  [ $\alpha = 0.0$ ] to  $\sim 85\%$  [ $\alpha = 1.0$ ] accuracy), and therefore may serve as the largest source of variability across masking studies using any type of mask with small  $\alpha$  values compared to studies using masks with larger  $\alpha$  values (at least for SOAs  $< 50$  ms). Therefore, such modulations of masking by the amplitude spectral slope could easily cloud any masking studies exploring scene gist masking caused by phase spectrum manipulations if the amplitude spectrum slopes vary extensively. Conversely, if amplitude spectrum slope is held approximately constant (as in Experiment 5 here), then it is possible to explore such phase spectrum effects. In doing so, we observed a target-mask category dissimilarity effect. One possible implication of such an effect might be that masks derived from wavelet techniques contain local image structure that can differentially mask the local features in target scenes in a target-mask category-dependent manner. Such feature masking may result in such masks interfering with mechanisms more directly linked to categorical processing.

It is worth noting that while we refer to two different masking effects in this study (i.e., the target-mask amplitude slope similarity principle for Experiments 1–3 and the target-mask categorical dissimilarity principle for Experiment 5), we do not mean to imply that they are independent from one another, or operate in any sort of hierarchical manner. Rather, we are simply emphasizing that specific Fourier characteristics can produce different masking effects. Specifically, the amplitude spectrum slope of a given mask plays a dominant role in the ability of that mask to disrupt rapid scene categorization, but it does so in a manner that is not contingent on the target-mask category relationship. Conversely, masks derived from wavelet techniques (used in Experiment 5) mask rapid scene categorization in a manner contingent on the target-mask category relationship. Thus, much caution must be taken when comparing masking functions across studies that employ masks with unrecognizable amplitude-only defined content versus unrecognizable conjunct amplitude + phase-defined image structures. Similarly, care must be taken in comparing masking

functions produced in studies employing masks derived from wavelet techniques versus those that do not. As the current study demonstrates, such important differences can lead to differential masking effects in rapid scene categorization tasks. The question of whether the two masking effects are independent of one another (or operate in some hierarchical manner) should be addressed in future studies in which, for example, the amplitude spectra slopes of wavelet-derived masks are systematically varied.

*Keywords:* real-world scenes, scene gist, rapid scene categorization, image statistics, amplitude spectrum slope, orientation bias, phase spectrum, visual backward masking, processing time, perceptual time course

## Acknowledgments

This research was supported in part by a Colgate Research Council grant to BCH and by grants from the Kansas State University Office of Sponsored Research and the NASA/Kansas Space Grant Consortium to LCL. We thank the two anonymous reviewers for helpful comments and Kelly Byrne and Ryan V. Ringer for assistance in experiment preparation and data collection. Parts of the research in this article were previously presented at the 2009 and 2012 Annual Meetings of the Vision Sciences Society, with their abstracts published in the *Journal of Vision*.

Commercial relationships: none.

Corresponding author: Bruce C. Hansen.

Email: bchansen@colgate.edu.

Address: Department of Psychology, Neuroscience Program, Colgate University, Hamilton, NY, USA.

## Footnote

<sup>1</sup>Cohen's  $f^2$  effect size values of 0.10, 0.25, and 0.40 are typically considered to be small, medium, and large effect sizes, respectively (Kirk, 1995).

## References

- Bacon-Mace, N., Mace, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*, 1459–1469.
- Brady, N., & Field, D. J. (1995). What's constant in contrast constancy? The effects of scaling on the

- perceived contrast of bandpass patterns. *Vision Research*, 35(6), 739–756.
- Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264(5163), 1333–1336.
- Carter, B. E., & Henning, G. B. (1971). The detection of gratings in narrow-band visual noise. *Journal of Physiology*, 219(2), 355–365.
- Cass, J., Alais, D., Spehar, B., & Bex, P. J. (2009). Temporal whitening: Transient noise perceptually equalizes the 1/f temporal amplitude spectrum. *Journal of Vision*, 9(10):12, 1019, <http://www.journalofvision.org/content/9/10/12>, doi:10.1167/9.10.12 [PubMed] [Article]
- Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences, USA*, 95(7), 4002–4006.
- De Valois, K. K., & Switkes, E. (1983). Simultaneous masking interactions between chromatic and luminance gratings. *Journal of the Optical Society of America*, 73(1), 11–18.
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22(5), 545–559.
- Essock, E. A., Haun, A. M., & Kim, Y.-J. (2009). An anisotropy of orientation-tuned suppression that matches the anisotropy of typical natural scenes. *Journal of Vision*, 9(1):35, 1–15, <http://www.journalofvision.org/content/9/1/35>, doi:10.1167/9.1.35. [PubMed] [Article]
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1):10, 1–29, <http://www.journalofvision.org/content/7/1/10>, doi:10.1167/7.1.10. [PubMed] [Article]
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, 4(12), 2379–2394.
- Field, D. J., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, 37(23), 3367–3383.
- Guyader, N., Chauvin, A., Peyrin, C., Héroult, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, 327, 313–318.
- Hansen, B. C., & Essock, E. A. (2004). A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision*, 4(12):5, 1044–1060, <http://www.journalofvision.org/content/4/12/5>, doi:10.1167/4.12.5. [PubMed] [Article]
- Hansen, B. C., Haun, A. M., & Essock, E. A. (2008). The “horizontal effect”: A perceptual anisotropy in visual processing of naturalistic broadband stimuli. In T. A. Portocello & R. B. Velloti (Eds.), *Visual cortex: New research*. New York, NY: Nova Science Publishers.
- Hansen, B. C., & Hess, R. F. (2007). Structural sparseness and spatial phase alignment in natural scenes. *Journal of the Optical Society of America A*, 24(7), 1873–1885.
- Hansen, B. C., & Hess, R. F. (2012). On the effectiveness of noise masks: Naturalistic vs. unnaturalistic image statistics. *Vision Research*, 60, 101–113.
- Heeger, D. J. (1992a). Half-squaring in responses of cat striate cells. *Visual Neuroscience*, 9(5), 427–443.
- Heeger, D. J. (1992b). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9(2), 181–197.
- Henning, G. B., Hertz, B. G., & Hinton, J. L. (1981). Effects of different hypothetical detection mechanisms on the shape of spatial-frequency filters inferred from masking experiments. I. Noise masks. *Journal of the Optical Society of America A*, 71, 574–581.
- Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 115–125.
- Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision*, 9(1):2, 1–16, <http://www.journalofvision.org/content/9/1/2>, doi:10.1167/9.1.2. [PubMed] [Article]
- Kaping, D., Tzvetanov, T., & Treue, S. (2007). Adaptation to statistical properties of visual scenes biases rapid categorization. *Visual Cognition*, 15(1), 12–19.
- Keil, M. S., & Cristobal, G. (2000). Separating the chaff from the wheat: Possible origins of the oblique effect. *Journal of the Optical Society of America A*, 17(4), 697–710.
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Legge, G. E., & Foley, J. M. (1980). Contrast masking

- in human vision. *Journal of the Optical Society of America*, 70(12), 1458–1471.
- Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(3), 435–441.
- Loftus, G. R., Hanna, A. M., & Lester, L. (1988). Conceptual masking: How one picture captures attention from another picture. *Cognitive Psychology*, 20(2), 237–282.
- Losada, M. A., & Mullen, K. T. (1995). Color and luminance spatial tuning estimated by noise masking in the absence of off-frequency looking. *Journal of the Optical Society of America A*, 12(2), 250–260.
- Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimari, T. (2010). The role of higher order image statistics in masking scene gist recognition. *Attention, Perception, & Psychophysics*, 72(2), 427–444.
- Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the natural/man-made distinction. *Journal of Vision*, 8(1):4, 1–9, <http://www.journalofvision.org/content/8/1/4>, doi:10.1167/8.1.4. [PubMed] [Article]
- Loschky, L. C., Sethi, A., Simons, D. J., Pydimari, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception & Performance*, 33(6), 1431–1450.
- Marr, D., Ullman, S., & Poggio, T. (1979). Bandpass channels, zero-crossings, and early visual information processing. *Journal of the Optical Society of America*, 69(6), 914–916.
- Michaels, C. F., & Turvey, M. T. (1979). Central sources of visual masking: Indexing structures supporting seeing at a single, brief glance. *Psychological Research-Psychologische Forschung*, 41(1), 1–61.
- Pavel, M., Sperling, G., Riedl, T., & Vanderbeek, A. (1987). Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America A*, 4(12), 2355–2365.
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1), 49–71.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, 2(5), 509–522.
- Rieger, J. W., Braun, C., Bulthoff, H. H., & Gegenfurtner, K. R. (2005). The dynamics of visual pattern masking in natural scene processing: A magnetoencephalography study. *Journal of Vision*, 5(3):10, 275–286, <http://www.journalofvision.org/content/5/3/10>, doi:10.1167/5.3.10. [PubMed] [Article]
- Sekuler, R. W. (1965). Spatial and temporal determinants of visual backward masking. *Journal of Experimental Psychology*, 70(4), 401–406.
- Solomon, J. A. (2000). Channel selection with non-white-noise masks. *Journal of the Optical Society of America A*, 17(6), 986–993.
- Stromeyer, C. F., & Julesz, B. (1972). Spatial-frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America A*, 62(10), 1221–1232.
- Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research*, 18(10), 1393–1399.
- Tolhurst, D. J., & Tadmor, Y. (2000). Discrimination of spectrally blended natural images: Optimisation of the human visual system for encoding natural images. *Perception*, 29(9), 1087–1100.
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, 14(3), 391–412.
- van der Schaaf, A., & van Hateren, J. H. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17), 2759–2770.
- VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Perception Science*, 2, 365.
- Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience*, 29, 10573–10581.
- Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vision Research*, 46(8-9), 1520–1529.
- Wilson, H. R., & Humanski, R. (1993). Spatial frequency adaptation and contrast gain control. *Vision Research*, 33(8), 1133–1149.
- Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23(9), 873–882.