Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions

Damon M. Chandler and Sheila S. Hemami

Visual Communications Laboratory, School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853

Received August 22, 2002; revised manuscript received January 6, 2003; accepted February 20, 2003

Quantization of the coefficients within a discrete wavelet transform subband gives rise to distortions in the reconstructed image that are localized in spatial frequency and orientation and are spatially correlated with the image. We investigated the detectability of these distortions. Contrast thresholds were measured for both simple and compound distortions presented in the unmasked paradigm and against two natural-image maskers. Simple and compound distortions were generated through uniform scalar quantization of one or two subbands. Unmasked detection thresholds for simple distortions yielded contrast sensitivity functions similar to those reported for 1-octave Gabor patches. Detection thresholds for simple distortions presented against two natural-image backgrounds revealed that thresholds were elevated across the frequency range of 1.15–18.4 cycles per degree with the greatest elevation for low-frequency distortions. Unmasked thresholds for compound distortions revealed relative sensitivities of 1.1–1.2, suggesting that summation of responses to wavelet distortions is similar to summation of responses to gratings. Masked thresholds for compound distortions revealed relative sensitivities of 1.5–1.7, suggesting greater summation when distortions are masked by natural images. © 2003 Optical Society of America


1. INTRODUCTION

Signal detection and discrimination have proved useful in characterizing many aspects of human vision, including spatial-frequency and orientation selectivity, channel bandwidths, gain control, and contour integration. In this paradigm, a signal (target) is presented against a usually well-defined background (mask), and thresholds indicate the observer’s ability to discriminate the signal + background from the background (or, equivalently, to detect the target in the presence of the mask).

Results of these types of psychophysical studies have been applied to a variety of image processing and compression algorithms, e.g., digital watermarking, texture analysis, Retinex-based image enhancement, and JPEG-2000. In particular, compression algorithms that aim to keep the compression-induced distortions below the threshold of detection rely heavily on established properties of visual sensitivity to those distortions; such algorithms are said to operate in a visually lossless manner. However, the application of psychophysics to image compression requires several key assumptions in order to validate the use of traditional psychophysical results in the image compression paradigm: (1) It is often assumed that visual sensitivity to traditional targets, such as sine waves and Gabor patches, is similar to sensitivity to compression-induced distortions (e.g., distortions induced through quantization of wavelet subbands), (2) it is often assumed that summation of visual responses to traditional targets is similar to summation of responses to algorithm-specific distortions, and (3) it is often assumed that results assessed in the unmasked paradigm are valid when targets (distortions) are presented against a natural image. Indeed, there is substantial psychophysical evidence asserting that visual processing of natural stimuli is unique.

This paper describes four psychophysical experiments designed to address these issues for targets consisting of wavelet subband quantization distortions, i.e., wavelet “textures” created through uniform scalar quantization of all coefficients within a wavelet subband (see Subsection 2.C). Experiment 1 measured unmasked detection thresholds for simple wavelet subband quantization distortions (i.e., distortions composed of 1 octave of spatial frequencies and a single orientation; see Fig. 5 below); this experiment was designed to assess differences between visual responses to wavelet distortions and responses to traditional targets. Experiment 2 measured detection thresholds for simple wavelet subband quantization distortions presented against two natural-image maskers; this experiment was designed to quantify the effects of natural-image backgrounds on detection thresholds and to provide insight into the types of distortions that are best masked by natural images. Experiment 3 measured unmasked detection thresholds for compound wavelet subband quantization distortions (i.e., distortions composed of 2 octaves of spatial frequencies or two orientations; see Fig. 10 below); this experiment was designed to investigate summation of visual responses to wavelet distortions on orientation and spatial-frequency dimensions. These results were compared with the results of experiment 4, which measured detection thresholds for compound wavelet distortions presented against two...
natural-image maskers; this latter experiment allowed a comparison of unmasked summation at threshold versus summation in the masked paradigm.

This paper is organized as follows: Section 2 provides a review of visual processing of natural images, visual summation, and wavelet subband quantization distortions. Section 3 describes the methods and the stimuli used in the experiments. Results and analyses are presented in Sections 4 and 5. A general discussion and an application to compression are presented in Section 6.

2. BACKGROUND

A. Visual Processing of Natural Images

It has been argued that natural images possess characteristic statistical regularities that have imposed evolutionary constraints on the functional role of visual processing units. For example, natural images exhibit characteristic amplitude spectra, which generally follow an $f^{-a}$ trend [where $f$ denotes spatial frequency; $a \in [0.7, 1.6]$ (Ref. 8)]; this property is believed to result from the scale-invariant and/or fractal nature of natural scenes.\(^9\)\(^10\)

Knill et al.\(^9\) have shown that human discrimination of fractal Brownian textures is optimal with $a \in [1.4, 1.8]$. Parraga et al.\(^11\) demonstrated that discrimination performance between morphed pairs of natural images was best for stimuli with $a = 1$. Using an adaptation paradigm, Webster and Miyahara\(^12\) have shown that changes in the slope of an image’s amplitude spectrum induces selective effects on contrast thresholds and suprathreshold contrast matches.

Natural images also possess a coherent phase structure, which is the primary contributor to an image’s phenomenal appearance. This fact was demonstrated by Oppenheim and Lim,\(^13\) who synthesized an image from the amplitude spectrum of one image and the phase spectrum of another; the resulting image appeared much more similar to the image whose phase structure was used. Thomson et al.\(^14\) have demonstrated that randomization or quantization of this phase structure severely impacts the semblance of an image. More recently, Bex and Makous\(^15\) have shown that randomizing a natural image’s phase structure at a particular spatial scale decreases detection and contrast-matching performance by the same amount as that caused by removing the spatial scale altogether. In addition, Geisler et al.\(^16\) have demonstrated that human performance in detecting contours can be predicted by means of a model based on the edge co-occurrence statistics of natural images.

The predominance of low spatial frequencies in natural images suggests slow changes in intensity, which are reflected by high positive interpixel correlations (i.e., neighboring pixels tend to have similar intensity values). Atick\(^17\) has argued that retinal ganglion cells may have evolved to remove these second-order (pairwise) correlations, yielding a “whitened” image as input to the lateral geniculate nucleus and area V1. Olshausen and Field\(^18\) have shown that training a neural network on natural images under a “sparse-coding” constraint yields a basis set that possesses similarities to cortical simple-cell receptive fields. Hyvärinen and Hoyer\(^19\) have shown similar correspondences between cortical complex-cell receptive fields and a basis set generated by training a multilayer neural network on natural images. Other nonlinear computational models have been used to demonstrate the phase-and shift-invariance properties of complex cells\(^20\) and higher-level effects such as end stopping and contour integration.\(^21\)

From an evolutionary standpoint, an organism must be efficient both at detecting a visual target within its natural environment and at blending in with this environment to avoid detection. This leads to the question: How effective are natural images at masking visual targets? Numerous models of visual masking have been quite successful at predicting detection thresholds for spatial targets placed on simple and complex backgrounds.\(^22\)\(^-\)\(^27\)

Many of these models have been developed and refined to fit various threshold-versus-contrast data,\(^22\) which generally indicate an increase in threshold as the contrast of the masker (pedestal) is increased and which often demonstrate a region of facilitation (i.e., a decrease in threshold; “dipper effect”) at lower pedestal contrasts, depending on the dimensional relationship between the target and the masker (e.g., differences in spatial frequency and orientation). The majority of these models employ a gain-control or “normalization” pool in which the combined response of several mechanisms is used to regulate the excitatory response of a detecting mechanism.\(^28\)\(^,\)\(^29\)

Indeed, variations of this framework have been incorporated into algorithms that have demonstrated success at predicting image discriminability\(^23\)\(^,\)\(^30\) and that have been successfully applied to image compression.\(^31\)\(^-\)\(^33\)

Still, it has proved difficult to quantify visual aspects of natural images, even on common dimensions (such as contrast)\(^15\)\(^,\)\(^34\) that are tractable for other stimuli. Part of this difficulty arises from the fact that natural images are compound stimuli whose components, together, constitute multiple spatial frequencies and multiple orientations; these components are thus subject to visual summation.

B. Summation of Visual Responses to Compound Targets

Although it is generally accepted that the human visual system (HVS) decomposes visual input through a bank of dimensionally localized channels,\(^35\) it is less clear how the responses of these channels are combined or “summed” to form what is ultimately seen. Summation experiments address this issue by comparing the detectability of a compound target with the detectability of its individual components (i.e., its components presented as simple targets). If the compound target is more easily detected than the simple targets, the visual responses to the compound target’s components are believed to have summed.\(^36\)

Let $t$ denote a simple target. The relative contrast of $t$, $RC(t)$, is defined as follows\(^36\):

$$RC(t) = \frac{C(t)}{CT(t)}, \quad (1)$$

where $C(t)$ is the contrast of $t$ and $CT(t)$ is the contrast threshold of $t$. Let $\tilde{t}$ denote a compound target composed
of two simple targets \(t_1\) and \(t_2\). The relative contrast threshold of \(t_i\), \(RCT(t_i; \tilde{t})\) (i = 1, 2), is defined as follows\(^36\):

\[
RCT(t_i; \tilde{t}) = \frac{\text{CT}(t_i|\tilde{t})}{\text{CT}(t_i)}, \tag{2}
\]

where \(\text{CT}(t_i|\tilde{t})\) is the contrast threshold of target \(t_i\) measured when \(t_i\) was presented as part of \(\tilde{t}\) and \(\text{CT}(t_i)\) is the contrast threshold of \(t_i\) measured when \(t_i\) was presented alone (i.e., as a simple target).

Note that when \(RCT(t_i; \tilde{t}) = 1\), \(\text{CT}(t_i|\tilde{t}) = \text{CT}(t_i)\), suggesting that the detectability of \(t_i\) is not affected by the presence of the other component \(t_j\) (j ≠ i) of \(\tilde{t}\). When \(RCT(t_i; \tilde{t}) < 1\), \(\text{CT}(t_i|\tilde{t}) < \text{CT}(t_i)\), suggesting that the detectability of \(t_i\) is enhanced by the presence of \(t_j\). When \(RCT(t_i; \tilde{t}) > 1\), \(\text{CT}(t_i|\tilde{t}) > \text{CT}(t_i)\), suggesting that the detectability of \(t_i\) is reduced by the presence of \(t_j\).

Summation is typically quantified through a single parameter denoting either a relative sensitivity\(^37\) (summation index, \(RS\) threshold ratio\(^38\)) or a Minkowski summation exponent (\(\beta\)). For a compound target composed of two components, the Minkowski sum (Quick approximation)\(^39\) is given by

\[
[RCT(t_1; \tilde{t})]^{1/\beta} + [RCT(t_2; \tilde{t})]^{1/\beta} = 1. \tag{3}
\]

When the components within the compound target are equally detectable, i.e., when \(RCT(t_1; \tilde{t}) = RCT(t_2; \tilde{t})\), relative sensitivity \(RS = 1/RCT(t_1; \tilde{t}) = 2^{1/\beta}\) (see Ref. 36). Thus, when \(\beta = 1\), \(RS = 2\), suggesting complete, or linear, summation (i.e., \(\frac{2}{2} + \frac{1}{2} = 1\) in Eq. (3)); whereas when \(\beta = \infty\), \(RS = 1\), suggesting no summation. In the latter case, the summation model functions as a maximum operator; i.e., the visual response to the compound is based only on the channel with the greatest output. Although we focus our current discussion on compound targets composed of two components, note that the Minkowski model was derived for the more general context in which compound targets are composed of \(N\) components (see, e.g., Ref. 40); in this case, when the components within the compound are equally detectable, \(RS = N^{1/\beta}\).

Summation at threshold on the orientation dimension was investigated by Carlson et al.\(^{41}\) and Manahilov and Simpson.\(^{37}\) Using a Yes–No procedure, Carlson et al. found \(RS = 1.18\) (\(\beta = 4.2\)) for compound targets composed of two sine-wave components oriented at 0 and 90 deg and \(RS = 1\) (\(\beta = \infty\)) when the components were oriented at 45 and 135 deg. Manahilov and Simpson found \(RS = 1.37\) (\(\beta = 2.2\)), using a two-alternative-forced-choice (2AFC) paradigm and compound targets composed of a pair of 6-cycle per degree (6-c/deg) Gabor patches oriented vertically and at 45 deg.

Summation at threshold on the spatial-frequency dimension was investigated by Graham and Nachmias,\(^{42}\) Sachs et al.,\(^{43}\) Watson,\(^{38}\) Manahilov and Simpson,\(^{37}\) and Meinhardt.\(^{44}\) Using the method of adjustment and a 2AFC paradigm, Graham and Nachmias found \(RS \approx 1\) for compound targets composed of two sine-wave components at spatial frequencies \(f_1\) and \(3f\); this result was reported to be invariant to the contrast ratio between the two components, invariant to the phase relationship between the two components, and invariant to the experimental paradigm (method of adjustment versus 2AFC procedure). Sachs et al. found similar results, using a Yes–No procedure and sine-wave targets composed of spatial frequencies \(f_1\) and \(f_2\); the components of the target were detected independently for most frequency ratios \(f_1/f_2 \in [0.80, 1.25]\). Using a 2AFC paradigm and Gabor targets, Watson found \(RS = 1.2\) (\(\beta = 3.6\)) for components separated by roughly 1 octave in spatial frequency. In line with the results from their investigation of summation on the orientation dimension, Manahilov and Simpson found \(RS = 1.37\) (\(\beta = 2.2\)) for a compound target composed of \((2 + 6)\)-c/deg Gabor patches. Meinhardt found slightly greater summation, \(RS = 1.44\) (\(\beta = 1.9\)), using the method of limits and sine-wave targets composed of spatial frequencies in the range \([1, 5]\) c/deg.

Using a 2AFC detection paradigm, Watson et al.\(^{45}\) measured summation at threshold on the spatial dimension. Contrast thresholds were measured for individual wavelet basis functions (simple targets) and for wavelet textures created by adding values drawn from a uniform distribution to an empty wavelet subband. Relative thresholds between these two stimuli yielded \(RS = 1.19\) (\(\beta \approx 4.0\)). Bonneh and Sagi\(^{40}\) measured masked summation on spatial-extent and spatial-position dimensions, using a compound target composed of 12.5-c/deg Gabor patches presented on a 30%–contrast 12.5-c/deg Gabor-patch pedestal. \(RS \in [1.20, 1.23]\) (\(\beta \in [3.3, 3.7]\)) was found when the extent of the pedestal was fixed at its maximal size or number; \(RS \in [1.16, 1.19]\) (\(\beta \in [4.0, 4.6]\)) was found on the

### Table 1. Summary of Results from Previous Summation Studies Using Far-Apart Targets

<table>
<thead>
<tr>
<th>Source</th>
<th>Dimension</th>
<th>Targets</th>
<th>Relative Sens.</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlson et al.(^{41})</td>
<td>Orientation</td>
<td>Sine waves</td>
<td>(\approx 1.0)</td>
<td>(\infty)</td>
</tr>
<tr>
<td>Manahilov and Simpson(^{37})</td>
<td>Orientation</td>
<td>Gabor</td>
<td>1.37</td>
<td>2.2</td>
</tr>
<tr>
<td>Graham and Nachmias(^{42})</td>
<td>Frequency</td>
<td>Sine waves</td>
<td>(\approx 1.0)</td>
<td>(\infty)</td>
</tr>
<tr>
<td>Sachs et al.(^{43})</td>
<td>Frequency</td>
<td>Sine waves</td>
<td>1.20</td>
<td>3.6</td>
</tr>
<tr>
<td>Watson(^{38})</td>
<td>Frequency</td>
<td>Gabor</td>
<td>1.37</td>
<td>2.2</td>
</tr>
<tr>
<td>Manahilov and Simpson(^{37})</td>
<td>Frequency</td>
<td>Gabor</td>
<td>1.44</td>
<td>1.9</td>
</tr>
<tr>
<td>Meinhardt(^{44})</td>
<td>Frequency</td>
<td>Gabor</td>
<td>1.19</td>
<td>(\approx 4.0)</td>
</tr>
<tr>
<td>Watson et al.(^{45})</td>
<td>Space</td>
<td>Wavelets</td>
<td>1.19</td>
<td>(\approx 4.0)</td>
</tr>
<tr>
<td>Bonneh and Sagi(^{40})</td>
<td>Space</td>
<td>Gabor</td>
<td>1.19</td>
<td>(\approx 4.0)</td>
</tr>
</tbody>
</table>
spatial-position dimension, and RS = 1 was found on the spatial-extent dimension when the extent of the pedestal was varied with that of the target.

In summary, numerous studies have investigated summation at threshold on the orientation, spatial-frequency, and spatial dimensions and masked summation on the spatial dimension. The results of these studies, which are summarized in Table 1, have revealed relative sensitivities ranging from 1.0 ($\beta = \infty$, no summation) to 1.44 ($\beta = 1.9$) with the use of sine-wave and Gabor targets presented against either a uniform background (no mask) or an unnatural masker. It remains unclear, however, whether these results are applicable to wavelet-based image compression in which the background is necessarily an image and the targets are wavelet distortions induced by quantization.

C. Wavelet Subband Quantization Distortions
State-of-the-art image compression algorithms attempt to mimic the multichannel nature of the HVS by employing a discrete wavelet transform (DWT) front end, which separates an image into spatial-frequency and orientation components. Although the DWT is not necessarily a good model of the decomposition performed during the early stages of human visual processing (cf. Ref. 46), the computational efficiency afforded by the DWT makes it particularly attractive for image compression and analysis. Specifically, the DWT is typically implemented by means of a filtering operation, usually in a separable fashion by successively filtering the rows and the columns of the image. This operation results in a tiling of the spatial-frequency plane, whereupon the image is represented as a series of spatial-frequency bands (called subbands). Figures 1 and 2 depict the frequency responses of the (one-dimensional) filters and the corresponding tiling of the spatial-frequency plane that results from a five-level DWT (by using the 9/7 biorthogonal filters\(^{45,47,48}\)).

As shown in Fig. 2, an N-level DWT will yield \(3N + 1\) subbands; each level contains an LH band, an HL band, and an HH band. The LH (HL) subbands are low-pass (high-pass) filtered horizontally and high-pass (low-pass) filtered vertically and hence contain horizontal (vertical) edge information with frequencies in the range given by the one-dimensional frequency response. Their corresponding spatial-frequency content can be described by their orientation and by their frequencies in cycles per pixel (which can be translated to cycles per degree once a viewing distance and a display resolution are established; see Ref. 45). The HH bands are high-pass filtered in both directions and thus contain both 45- and 135-deg edge information. The three subbands in a level are collectively referred to as a scale, with finer scales containing higher-frequency information and coarser scales containing lower-frequency information. The coarsest scale contains four subbands, in which the additional band is the low-frequency band (represented by the lower left of the tiling shown in Fig. 2).

After an image is transformed into its spatial-frequency representation, the coefficients within each subband are quantized. Quantization is a noninvertible process in which a continuous set of input values (e.g., subband coefficients) is approximated by a discrete set of output levels (called reproduction values). Let \(c(s)\) denote a coefficient of subband \(s\), and let \(C = \{C_n\}\) denote a partition of the real line into contiguous, nonoverlapping intervals \((C_n)\). A scalar quantizer operates by mapping each coefficient \(c(s) \in C_n\) to the reproduction value \(c_n(s)\). The width of the interval \(C_n\) is called the quantizer step size, denoted \(\Delta_n\). If the intervals are equispaced (i.e., \(\Delta_n = \Delta \forall n\)) and if the reproduction values \([c_n(s)]\) are midway between adjacent intervals, the quantizer is said to be uniform with a single quantizer step size \(\Delta^{49}\).

Quantization is modeled by the addition of noise to the original image. Specifically, quantization of a subband coefficient \(c(s)\) induces an error \(d(s) = c_n(s) - c(s)\), which manifests itself in the reconstructed image as a wavelet micropattern (distortion) whose amplitude is proportional to \(d(s)\|\Psi(s)\|\), where \(\Psi(s)\) represents the wavelet basis function associated with subband \(s\). When all coefficients of subband \(s\) are quantized, the resulting distortions constitute a superposition of wavelet micropatterns (i.e., a wavelet texture). Thus the reconstructed image \(\hat{I} = I + D_{s,\Delta}\), where \(I\) denotes the original image and \(D_{s,\Delta}\) denotes the wavelet distortions induced through uniform scalar quantization of subband \(s\) with step size \(\Delta\); this process is illustrated in Fig. 3. Note in Fig. 3 that the distortions are spatially correlated with the image.

Previous studies have measured human visual sensitivities to simple wavelet subband quantization distortions, i.e., distortions induced through quantization of single DWT subbands. In an unmasked detection experiment, Watson et al.\(^{45}\) used a 2AFC paradigm to measure...
the visibility of both individual wavelet basis functions
and noise injected into individual DWT subbands. (Note
that the distortions produced by this latter technique do
not possess spatial correlations that are induced through
actual quantization; cf. Fig. 3). Ramos and Hemami48
performed a similar experiment in the masked paradigm
by using the method of adjustment and stimuli created
through actual quantization of DWT subbands obtained
from various natural-image maskers. These studies
yielded at-threshold and suprathreshold quantizer step
sizes for individual subbands, which are directly appli-
cable when only one subband is quantized, i.e., when the
quantization distortions encompass an octave-limited
band of frequencies. However, in a compression applica-
tion, all $3N + 1$ subbands of an $N$-level DWT are quan-
tized simultaneously, which requires an adjustment of the
individual quantizer step sizes such that the resulting
compound distortions meet the desired detectability crite-
rion. In the spirit of previous summation-at-threshold
experiments, to produce images with no visible distortion,
Watson et al.45 assumed that the combined distortions
were negligibly more detectable than the individual dis-
tortions (i.e., $\beta = \infty$). To produce highly compressed, vi-
sually lossy images, Ramos and Hemami48 assumed that
responses to quantization distortions pooled linearly at
suprathreshold contrasts and accordingly used a
Minkowski metric with $\beta = 1$. Although these assump-
tions were made for different contrast ranges, it remains
unclear which model is better suited to compressing natu-
ral images. Section 3 describes experiments that were
performed to address this issue.

3. METHODS

Four experiments were conducted to quantify the effects
of natural images on the detectability of wavelet subband
quantization distortions. In experiment 1, unmasked de-
tection thresholds were measured for simple wavelet dis-
tortions. In experiment 2, detection thresholds were
measured for these same distortions presented against
two different natural-image maskers. In experiment 3,
unmasked detection thresholds were measured for com-
 pound wavelet distortions composed of either 2 octaves of
spatial frequencies or two orientations. In experiment 4,
detection thresholds were measured for these compound
distortions presented against the same image maskers as
those of experiment 2.

A. Apparatus

Stimuli were displayed on a high-resolution, noninter-
laced HP A4033A 19-in. monitor (0.26-mm dot pitch, 82-
kHz horizontal frequency, and 120-Hz vertical frequency)
at a display resolution of 36.4 pixels/cm, a frame rate of
75 Hz, and an overall gamma of 2.3. The display yielded
minimum, maximum, and mean luminances of, respec-
tively, 0.08, 48.2, and 13.3 cd/m$^2$. Stimuli were viewed
binocularly through natural pupils in a darkened room at
a distance of approximately 58 cm, resulting in a display
visual resolution of 36.8 pixels/deg.

B. Stimuli

Stimuli consisted of luminance modulations of size
$512 \times 512$ pixels (13.9 deg $\times$ 13.9 deg). Each stimulus
was composed of a target and a mask: In all experi-
ments, targets consisted of wavelet subband quantization
distortions; in experiments 1 and 3 (unmasked detection),
the mask consisted of a uniform gray image (10.1 cd/m$^2$);
in experiments 2 and 4, the mask consisted of a natural
image.

Wavelet targets were generated through uniform scalar
quantization of one or two DWT subbands, yielding
simple or compound distortions, respectively; the former
were used in experiments 1 and 2, and the latter were
used in experiments 3 and 4. The subbands were
obtained by transforming a natural image of size
$512 \times 512$ pixels using five decomposition levels and the
9/7 biorthogonal DWT filters (also used by Watson et al.45
and Ramos and Hemami48; see also Ref. 47). At the dis-
play visual resolution of 36.8 pixels/deg, the LH and HL
subbands at the first through fifth levels correspond to
center spatial frequencies of, respectively, 18.4, 9.2, 4.6,
2.3, and 1.15 c/deg. We acknowledge that this range of
frequencies is relatively limited as compared with that in previous studies, which have measured contrast sensitivity to gratings. Sensitivity to lower spatial frequencies could be tested by using more decomposition levels; sensitivity to higher spatial frequencies could be tested by increasing the display visual resolution.

Simple targets. Simple horizontal targets were generated by uniformly quantizing an LH subband of DWT level 1, 2, 3, 4, or 5, yielding a superposition of horizontally oriented wavelet basis functions centered at spatial frequency 18.4, 9.2, 4.6, 2.3, or 1.15 c/deg, respectively. Simple vertical targets were generated in a similar fashion through quantization of the HL subbands. The quantizer step size for each subband was selected such that the rms contrast of the resulting distortions was as requested by the adaptive staircase procedure described in Subsection 4.C.

Compound targets. Compound targets composed of two orientations were generated by uniformly quantizing the LH and HL subbands of DWT level 3, 4, or 5, yielding a superposition of horizontally and vertically oriented targets centered at spatial frequency 4.6, 2.3, or 1.15 c/deg, respectively. Horizontally oriented compound targets composed of two spatial frequencies were generated by uniformly quantizing the LH subbands of DWT level 4 and 3 or 5 and 4, yielding a superposition of targets centered at spatial frequency 2.3 + 4.6 or 1.15 + 2.3 c/deg, respectively. Vertically oriented compound targets composed of the same pairs of spatial frequencies were generated in a similar fashion through uniform quantization of the HL subbands. For each compound target, the relative contrasts of its components were equalized by using an estimate of each simple target’s threshold (obtained from a previous study using the same images; see Ref. 50). Quantizer step-size pairs were selected to meet this relative-contrast criterion and such that the rms contrast of the compound target was as requested by the adaptive staircase procedure described in Subsection 4.C.

Following quantization of the subband(s), an inverse DWT was applied to generate a reconstructed image (target + mask) of size 512 × 512 pixels. For experiments 1 and 3, the mask (image) was subtracted from the reconstructed image, yielding only the target (wavelet distortions); the target was then added to an equally sized uniform gray image. This technique allowed unmasked presentation of targets while preserving spatial correlations between the distortions and the natural images (see Fig. 3).

Two natural images, a balloon and a horse, were used in all experiments as sources of the distortions and as maskers (for experiments 2 and 4); these images are depicted in Fig. 4. Both images were of size 512 × 512 pixels and were 8-bit gray scale with pixel values in the range 0–255. The displayed images had mean luminances of 15.7 cd/m² (balloon) and 10.9 cd/m² (horse).

C. Experimental Procedures

Thresholds were measured by using a spatial three-alternative forced-choice procedure. On each trial, observers concurrently viewed three adjacent images placed on a uniform 10.1 cd/m² background. Two of the images contained the mask alone, and the other image additionally contained one of the previously described targets (distortions). Observers indicated by means of keyboard input which of the three images contained the target. Target contrasts were controlled through a QUEST staircase procedure using software derived from the Psychophysics Toolbox. Contrast threshold was defined as the 75%-correct point on a Weibull function, which was fitted to the data following each series of trials.

Each experimental session began with at least two blocks of 32 practice trials following 3 min each of dark adaptation and adaptation to a uniform 10.1 cd/m² display. Before each series of trials, observers were briefly shown a high-contrast, spatially randomized version of the distortions to minimize subjects’ uncertainty in the orientation and the frequency of the target. During each trial, an auditory tone indicated stimulus onset, and auditory feedback was provided on an incorrect response. The image to which the target was added was randomly selected at the beginning of each trial, and observers were instructed to examine all three images before responding. Response time was limited to within 7 s of stimulus onset, during which all three images remained visible.

Each series of trials consisted of two interleaved tracks of 32-trial blocks, one track for each source-image condition (balloon and horse). Thus a different masker and/or

Fig. 4. Two 512 × 512 natural images (balloon and horse) used as masks in this study.
source of distortions was used on every trial; however, the orientation and the spatial frequency of the target (or combinations of these parameters for compound targets) remained constant throughout the entire block, and this same orientation/frequency combination was used for both staircase tracks (image conditions); i.e., observers knew to look for a target of a specific orientation and center spatial frequency, regardless of the image masker/source. We acknowledge that learning and light and pattern adaptation had likely occurred during the course of a block; however, aside from limiting response time and interleaving conditions, we did not attempt to control for these factors, as they are characteristic of natural viewing conditions.

D. Observers
The first author (DC) and two naive adult subjects (MM and SC) participated in the experiments. All observers were familiar with compression-induced distortions; however, only DC had previous exposure to the image maskers. All had either normal or corrected-to-normal visual acuity.

E. Contrast Metric
Results are reported here in terms of rms contrast, a metric that has proved useful for threshold measurements using compound, noise, wavelet, and natural-image stimuli. The rms contrast of a stimulus is computed as follows:

\[ C_{\text{rms}} = \left( \frac{1}{L_{\text{mask}} \cdot \sum_{i=0}^{N} (L_i - L_{\text{mask}})^2} \right)^{1/2}, \]  

where \( C_{\text{rms}} \) denotes the rms contrast, \( L_{\text{mask}} \) the average masker luminance, \( L_i \) the luminance of the \( i \)th pixel, and \( N \) the total number of pixels. Note that, in this study, because observers’ fixations were not tracked, spatially localized contrast is not considered.

4. DETECTION OF SIMPLE WAVELET DISTORTIONS: RESULTS AND ANALYSIS

A. Unmasked Detection of Simple Wavelet Targets
Experiment 1 measured unmasked detection thresholds for simple horizontal and vertical wavelet subband quantization distortions at center spatial frequencies of 1.15, 2.3, 4.6, 9.2, and 18.4 c/deg. On each trial, observers were presented with three images: two uniform gray images (10.1 cd/m²) and a uniform gray image to which distortions were added as described in Subsection 3.B. Figure 5 depicts suprathreshold versions of the stimuli used in experiment 1.

Thresholds for each observer are plotted in Fig. 6 as a function of the center spatial frequency of the distortions. For the five spatial frequencies and the two orientations tested, these data indicate the minimum rms contrast necessary to detect the target (distortions) in the presence of no mask. Each data point in Fig. 6 represents the average of at least two blocks of trials. The error bar for each point denotes the standard error of the mean (SE).

The results of experiment 1 are generally consistent with those of previous contrast threshold measurements; namely, similar to what has been found for gratings, the minimum contrast required to detect wavelet subband quantization distortions varies with spatial frequency, and equal sensitivity is observed for horizontal and vertical distortions.

Contrast sensitivity to sine-wave targets is traditionally found to peak at 2–6 c/deg. In general, the data of Fig. 6 exhibit a maximum at 1.15 c/deg. These variations are perhaps attributable, in part, to differences in the bandwidth of sine-wave gratings and wavelet subband quantization distortions. Specifically, as opposed to sine-wave gratings, which occupy a single point in frequency, wavelet subband quantization distortions encompass 1 octave of spatial frequencies. Although current evidence (including results presented in Subsection 5.A) indicates that HVS spatial-frequency channels have a tuning bandwidth of approximately 1 octave, other studies employing stimuli with bandwidths near 1 octave have also found low-pass-shaped contrast sensitivity functions (CSFs). The results of Peli et al., for example, revealed maximum contrast sensitivity at 1–2 c/deg for 1-octave Gabor patches. Thresholds measured for wavelet subband quantization distortions by Watson et al. also revealed maximum sensitivity at the lowest spatial frequency tested (2 c/deg).

B. Masked Detection of Simple Wavelet Targets
Experiment 2 measured detection thresholds for simple horizontal and vertical wavelet subband quantization distortions at center spatial frequencies 1.15, 2.3, 4.6, 9.2, and 18.4 c/deg presented against two different natural-image maskers (balloon and horse). On each trial, observers were presented with three images: two copies of the masker and one of the stimuli (mask + simple target), as described in Subsection 3.B. Figure 7 depicts suprathreshold versions of the stimuli used in experiment 2.

Figure 8 depicts the masked thresholds obtained for each observer as a function of the spatial frequency of the target. For the five spatial frequencies and the two orientations tested, these data indicate the minimum rms contrast necessary to detect the target in the presence of the corresponding natural-image masker. Each data point in Fig. 8 represents the average of at least two blocks of trials; error bars denote SEs.

The results depicted in Fig. 8 provide insight into the types of quantization distortions that are readily masked by natural images. Namely, whereas in experiment 1 maximum sensitivity occurred at 1.15 c/deg, these data show approximately equal sensitivity to 1.15–2.3–, and 4.6-c/deg targets. These differences in maximum sensitivity between the data of Figs. 6 and 8 may be attributed, in part, to the characteristic amplitude spectra of natural images as described in Subsection 2.A. In particular, the lower frequencies that predominate natural images might also reduce the visibility of lower-frequency distortions. This notion is illustrated in Fig. 9, which depicts contrast threshold elevations (TEs) between the masked and unmasked conditions (i.e., \( C_{\text{T masked}} / C_{\text{T unmasked}} \)). Previous spatial masking experiments have traditionally found...
greatest elevations in thresholds when the spatial frequency of the mask is near that of the target. When the mask is a natural image, which is composed predominantly of low spatial frequencies, the greatest elevation in threshold occurs for low-spatial-frequency distortions (TE = 8 to 9 at 1.15 c/deg for DC), whereas distortions of
5. DETECTION OF COMPOUND WAVELET DISTORTIONS: RESULTS AND ANALYSIS

A. Unmasked Detection of Compound Wavelet Targets

Experiment 3 measured unmasked detection thresholds for compound wavelet targets composed either of two orientations (horizontal + vertical), both at center spatial frequency 1.15, 2.3, or 4.6 c/deg, or 2 octaves of spatial frequencies (2.3 + 4.6 and 1.15 + 2.3 c/deg), both oriented horizontally or vertically. On each trial, observers were presented with three images: two uniform gray images (10.1 cd/m²) and a uniform gray image to which the compound target was added, as described in Subsection 3.B. Figure 10 depicts suprathreshold versions of the stimuli used in experiment 3.

Results of this experiment were compared with those of experiment 1 to quantify summation of visual responses to wavelet targets on orientation and spatial-frequency dimensions in the unmasked paradigm. Namely, if responses to the individual components of a compound wavelet target exhibit summation, then a reduction in detection thresholds for the components within the compound target (measured in experiment 3) versus threshold (TE ~ 2 to 3 at 18.4 c/deg for DC).

high spatial frequencies incur only a minor elevation in threshold (TE ~ 2 to 3 at 18.4 c/deg for DC).
olds measured for those same components presented individually (experiment 1) would be expected. With the notation presented in Subsection 2.B, the $CT(t_i)$ were measured in experiment 1 (see Subsection 4.A) and the $CT(t_i \parallel t_f)$ were measured in experiment 3.

Figures 11 and 12 depict relative contrast thresholds of experiment 3 in the form of summation-square plots (see Fig. 13). In each plot, the horizontal axis represents the relative contrast of one component [i.e., $RC(t_1)$], and the vertical axis represents the relative contrast of the other component [i.e., $RC(t_2)$]. The points within each plot denote relative contrast threshold pairs $(RCT(t_1; t_f), RCT(t_2; t_f))$ computed by using each observer’s average contrast threshold for each component (from experiment 1); each data point represents the average of at least two blocks of trials. Thus, for linear summation ($RS = 2, \beta = 1$), the data points would lie on the diagonal line connecting $RC$ coordinates $(0,1)$ to $(1,0)$; for no summation ($RS = 1, \beta = \infty$), the points would lie on the lines formed by connecting $(0,1)$ to $(1,1)$ and $(1,1)$ to $(1,0)$. 

![Fig. 9. Contrast threshold elevations (masked/unmasked) imposed by each natural image on the detectability of wavelet distortions. Black circles, data for horizontal targets; gray circles, data for vertical targets.](image)

![Fig. 10. Representative stimuli used in experiment 3: compound wavelet subband quantization distortions composed of horizontal \(\parallel\) vertical components at center frequencies (a) 2.3 c/deg and (b) 1.15 c/deg and compound wavelet subband quantization distortions composed of 2 octaves of frequencies centered at 1.15 + 2.3 c/deg oriented (c) horizontally and (d) vertically. Distortions were generated by quantizing subbands from the horse image. Stimuli containing quantization distortions at center frequencies 4.6, 9.2, and 18.4 c/deg, and distortions generated from the balloon image, are not depicted.](image)

![Fig. 11. Relative contrast thresholds measured in the unmasked paradigm (experiment 3) for compound wavelet subband quantization distortions containing orthogonal components of equal spatial frequencies. Distortions were generated by quantizing the LH and HL subbands of the balloon and horse images. The horizontal and vertical axes of each plot represent the relative rms contrasts of the horizontal and vertical components, respectively. Open circles, 4.6-c/deg components; gray circles, 2.3-c/deg components; black circles, 1.15-c/deg components. Solid curves indicate fits of Eq. (3) to the data (one curve for each data point); each curve is color coded and corresponds to the data point through which it runs.](image)
Figure 11 depicts relative contrast thresholds for compound targets composed of orthogonal components of equal spatial frequencies. Figure 12 depicts relative contrast thresholds for compound targets composed of equally oriented components of different center spatial frequencies. The solid curves in each plot represent fits of Eq. (3) to each data point.

Relative sensitivities computed from the data of Figs. 11 and 12 are listed in Tables 2 and 3. Results for summation on the orientation dimension yield mean\textsuperscript{38} relative sensitivities (over all three observers) of 1.21 (\(\beta = 3.64\)) and 1.15 (\(\beta = 5.06\)) for targets generated from the balloon and horse images, respectively, with an overall mean (over both images) of 1.19 (\(\beta = 3.97\)). Results for summation on the spatial-frequency dimension yield mean relative sensitivities of 1.15 (\(\beta = 5.01\)) and 1.13 (\(\beta = 5.62\)) for targets generated from the balloon and horse images, respectively, with an overall mean of 1.14 (\(\beta = 5.30\)).

These results are consistent with previous summation experiments\textsuperscript{37,38,41,43,44,61} indicating that, in the unmasked paradigm, HVS responses to wavelet subband quantization distortions pool in a fashion similar to that of responses to sine-wave gratings and Gabor patches. In particular, when the components of the compound target are far enough apart along the dimension of interest (1 octave; 90 deg in this experiment), doubling the number of components does not halve the contrasts of the components when the compound is at threshold; rather, the contrast of each component is reduced only by a scaling factor of approximately 0.85 (corresponding to \(\beta \approx 4\) to 5). These data provide further evidence that HVS orientation and spatial-frequency channels have approximately 1-octave tuning bandwidths.

In addition, these results are consistent with those from a previous summation study\textsuperscript{50} in which compound targets generated from the balloon and horse images were analyzed. Table 2 presents the relative sensitivities for unmasked wavelet targets of the balloon image, while Table 3 provides similar data for the horse image.

![Figure 12](image-url)

Figure 12. Relative contrast thresholds measured in the unmasked paradigm (experiment 3) for compound wavelet subband quantization distortions containing equally oriented components of different center spatial frequencies. Distortions were generated by quantizing the LH\(_n\) (HL\(_n\)) and LH\(_{n+1}\) (HL\(_{n+1}\)) subbands (\(n = 3, 4\)) of the balloon and horse images. The horizontal and vertical axes represent the relative rms contrasts of the higher-frequency and lower-frequency components, respectively. Black circles, horizontal 2.3 + 4.6 c/deg; gray circles, vertical 2.3 + 4.6 c/deg; black squares, horizontal 1.15 + 2.3 c/deg; gray squares, vertical 1.15 + 2.3 c/deg. Solid curves indicate fits of Eq. (3) to the data (one curve for each data point); each curve is color coded and corresponds to the data point through which it runs.

![Figure 13](image-url)

Figure 13. Example of a summation-square plot denoting regions of linear summation, probability summation, and no summation. The horizontal axis corresponds to the relative contrast of one of the compound target’s components; the vertical axis corresponds to the relative contrast of the other component. For linear (complete) summation (RS = 2, \(\beta = 1\)), relative contrast thresholds would fall on the diagonal line connecting coordinates (0,1) to (1,0). For no summation (RS = 1, \(\beta = \infty\)), relative contrast thresholds would fall on the lines formed by connecting (0,1) to (1,1) and (1,1) to (1,0). The majority of summation-at-threshold experiments have found relative contrast thresholds to lie between these two extremes, typically with RS \(\approx 1.2\) (probability summation, see Ref. 36).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Components</th>
<th>Relative Sens.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>H+V 1.15 c/deg</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.27</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.45</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>H 1.15+2.3 c/deg</td>
<td>1.14</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>V 1.15+2.3 c/deg</td>
<td>1.27</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>H 2.3+4.6 c/deg</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>V 2.3+4.6 c/deg</td>
<td>1.00</td>
<td>0.10</td>
</tr>
<tr>
<td>MM</td>
<td>H+V 1.15 c/deg</td>
<td>1.00</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.24</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.33</td>
<td>0.06</td>
</tr>
<tr>
<td>SC</td>
<td>H+V 1.15 c/deg</td>
<td>1.17</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.26</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.31</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>H 1.15+2.3 c/deg</td>
<td>1.08</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>V 1.15+2.3 c/deg</td>
<td>1.34</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>H 2.3+4.6 c/deg</td>
<td>1.35</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>V 2.3+4.6 c/deg</td>
<td>1.30</td>
<td>0.10</td>
</tr>
</tbody>
</table>
wavelet subband distortions were generated by adding noise into pairs of DWT subbands (rather than through actual quantization). Whereas actual quantization-induced distortions exhibit spatial correlations with the source image, noise added to a DWT subband results in distortions that are uniformly distributed throughout the stimulus. Our current results therefore suggest that summation at threshold on the spatial-frequency and orientation dimensions does not depend on the spatial configuration of the targets.

### B. Masked Detection of Compound Wavelet Targets

Experiment 4 measured masked thresholds for the same compound wavelet targets as those of experiment 3 (see Subsection 5.A) using the balloon and horse natural-image maskers. On each trial, observers were presented with three images: two copies of the masker and one of the stimuli (mask + compound target), as described in Subsection 3.B. Figure 14 depicts suprathreshold versions of the stimuli used in experiment 4.

**Table 3. Relative Sensitivities for Unmasked Wavelet Targets of the Horse Image**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Components</th>
<th>Relative Sens.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>H+V 1.15 c/deg</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.23</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.18</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>H 1.15+2.3 c/deg</td>
<td>1.17</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>V 1.15+2.3 c/deg</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>H 2.3+4.6 c/deg</td>
<td>1.29</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>V 2.3+4.6 c/deg</td>
<td>1.08</td>
<td>0.00</td>
</tr>
<tr>
<td>MM</td>
<td>H+V 1.15 c/deg</td>
<td>1.16</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.11</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.19</td>
<td>0.00</td>
</tr>
<tr>
<td>SC</td>
<td>H+V 1.15 c/deg</td>
<td>1.29</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.13</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.08</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>H 1.15+2.3 c/deg</td>
<td>1.08</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>V 1.15+2.3 c/deg</td>
<td>1.33</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>H 2.3+4.6 c/deg</td>
<td>1.00</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>V 2.3+4.6 c/deg</td>
<td>1.20</td>
<td>0.02</td>
</tr>
</tbody>
</table>

**Fig. 14.** Representative stimuli used in experiment 4: horse image containing compound wavelet subband quantization distortions composed of horizontal + vertical components at center frequencies (a) 2.3 c/deg and (b) 1.15 c/deg; horse image containing compound wavelet subband quantization distortions composed of 2 octaves of frequencies centered at 1.15 + 2.3 c/deg oriented (c) horizontally and (d) vertically. Stimuli containing quantization distortions at center frequencies 4.6, 9.2, and 18.4 c/deg, and stimuli containing the balloon image, are not depicted.

**Fig. 15.** Relative contrast thresholds measured in the masked paradigm (experiment 4) for compound wavelet subband quantization distortions containing orthogonal components of equal spatial frequencies. Distortions were generated by quantizing the LH and HL subbands of the natural images described in Subsection 3.B. The horizontal and vertical axes of each plot represent the relative rms contrasts of the horizontal and vertical components, respectively. Open circles, 4.6-c/deg components; gray circles, 2.3-c/deg components; black circles, 1.15-c/deg components. Solid curves indicate fits of Eq. (3) to the data (one curve for each data point); each curve is color coded and corresponds to the data point through which it runs.

B. Masked Detection of Compound Wavelet Targets

Experiment 4 measured masked thresholds for the same compound wavelet targets as those of experiment 3 (see Subsection 5.A) using the balloon and horse natural-image maskers. On each trial, observers were presented with three images: two copies of the masker and one of the stimuli (mask + compound target), as described in Subsection 3.B. Figure 14 depicts suprathreshold versions of the stimuli used in experiment 4.

Thresholds measured in this experiment were compared with those from experiment 2 to quantify summation of visual responses to wavelet distortions presented against natural-image maskers on orientation and spatial-frequency dimensions. Figures 15 and 16 depict the relative contrast thresholds. Relative contrast thresholds for compound targets composed of orthogonal components of equal spatial frequencies are shown in Fig.
quantizing the LH
ferent center spatial frequencies. Distortions were generated by
paradigm (experiment 4) for compound wavelet subband quanti-
Fig. 16. Relative contrast thresholds measured in the masked

ter spatial frequencies are shown in Fig. 16. The hori-

zontal and vertical axes each represent the relative con-
trast of a component within the compound target. Data
points denote relative contrast threshold pairs
(RCT(t1; t), RCT(t2; t)) computed by using each observ-
er’s average contrast threshold for each component (from
experiment 2). Each data point represents the average
of at least two blocks of trials; solid curves represent fits
of Eq. (3) to each point.

Relative sensitivities computed from the data of Figs.
15 and 16 are listed in Tables 4 and 5. Results for sum-
mation on the orientation dimension yield mean
relative sensitivities (over all three observers) of 1.53
(β = 1.62) and 1.56 (β = 1.57) for targets generated
from the balloon and horse images, respectively, with an
overall mean (over both images) of 1.54 (β = 1.60).
Results for summation on the spatial-frequency dimension
yield mean relative sensitivities of 1.71 (β = 1.29) and
1.65 (β = 1.39) for targets generated from the balloon
and horse images, respectively, with an overall mean of
1.68 (β = 1.34).

These data indicate that summation of responses to
wavelet subband quantization distortions presented
against a natural-image masker is significantly greater
than what was found for these same distortions in the un-
masked paradigm (cf. Tables 2 and 3). These high rela-
tive sensitivities are perhaps attributable to within-
channel summation, which may arise because of the
spatial correlations that exist between the distortions and
the image maskers. Similarly, increased summation
might be attributable to “off-frequency looking,”62,63
whereby the channel with the greatest signal-to-noise ra-
tio, though less optimally tuned to the target, is used for
detection. Thus, assuming that the image masker lowers
the signal-to-noise ratio of the channel tuned to one com-
ponent, the other channel (i.e., the channel optimal for
the other component and “switched to” for the first com-
ponent) might elicit detection for both components, again
giving rise to within-channel summation.

Table 4. Relative Sensitivities for Masked Wavelet
Targets of the Balloon Image

<table>
<thead>
<tr>
<th>Subject</th>
<th>Components</th>
<th>Relative Sens.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>H+V 1.15 c/deg</td>
<td>1.68</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.63</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.59</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>H 1.15+2.3 c/deg</td>
<td>1.45</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>V 1.15+2.3 c/deg</td>
<td>1.50</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>H 2.3+4.6 c/deg</td>
<td>2.25</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>V 2.3+4.6 c/deg</td>
<td>2.09</td>
<td>0.27</td>
</tr>
<tr>
<td>MM</td>
<td>H+V 1.15 c/deg</td>
<td>1.36</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.42</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.42</td>
<td>0.48</td>
</tr>
<tr>
<td>SC</td>
<td>H+V 1.15 c/deg</td>
<td>1.71</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.36</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.76</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>H 1.15+2.3 c/deg</td>
<td>1.66</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>V 1.15+2.3 c/deg</td>
<td>1.69</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>H 2.3+4.6 c/deg</td>
<td>1.77</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>V 2.3+4.6 c/deg</td>
<td>1.55</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 5. Relative Sensitivities for Masked Wavelet
Targets of the Horse Image

<table>
<thead>
<tr>
<th>Subject</th>
<th>Components</th>
<th>Relative Sens.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>H+V 1.15 c/deg</td>
<td>1.92</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.35</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.65</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>H 1.15+2.3 c/deg</td>
<td>1.64</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>V 1.15+2.3 c/deg</td>
<td>1.90</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>H 2.3+4.6 c/deg</td>
<td>1.78</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>V 2.3+4.6 c/deg</td>
<td>1.73</td>
<td>0.36</td>
</tr>
<tr>
<td>MM</td>
<td>H+V 1.15 c/deg</td>
<td>1.63</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.60</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.81</td>
<td>0.64</td>
</tr>
<tr>
<td>SC</td>
<td>H+V 1.15 c/deg</td>
<td>1.72</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>H+V 2.3 c/deg</td>
<td>1.26</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>H+V 4.6 c/deg</td>
<td>1.33</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>H 1.15+2.3 c/deg</td>
<td>1.86</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>V 1.15+2.3 c/deg</td>
<td>1.27</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>H 2.3+4.6 c/deg</td>
<td>1.47</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>V 2.3+4.6 c/deg</td>
<td>1.73</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Fig. 16. Relative contrast thresholds measured in the masked
paradigm (experiment 4) for compound wavelet subband quanti-

tions containing equally oriented components of dif-
ferent center spatial frequencies. Distortions were generated by
quantizing the LHn (HLn) and LHn+1 (HLn+1) subbands
(n = 3, 4) of the natural images described in Subsection 3.B.
The horizontal and vertical axes represent the relative rms con-
trasts of the higher-frequency and lower-frequency components,
respectively. Black circles, horizontal 2.3 + 4.6 c/deg; gray
circles, vertical 2.3 + 4.6 c/deg; black squares, horizontal
1.15 + 2.3 c/deg; gray squares, vertical 1.15 + 2.3 c/deg. Solid
curves indicate fits of Eq. (3) to the data (one curve for each data
point); each curve is color coded and corresponds to the data
point through which it runs.
6. DISCUSSION

Previous studies have investigated the detectability of simple and compound stimuli presented against no mask or against well-defined maskers. In this study, we have examined the detectability of simple and compound wavelet subband quantization distortions presented against no mask and two natural-image maskers. Detection thresholds measured in the unmasked paradigm were compared with detection thresholds measured when the distortions were presented against two natural-image backgrounds.

A. Detection of Simple Wavelet Distortions

Contrast sensitivity functions (CSFs) for sine-wave gratings traditionally demonstrate a bandpass profile, with a peak at 2–6 c/deg. CSFs for Gabor patches usually demonstrate a low-pass profile, peaking at 0.5–3 c/deg, depending on the bandwidth of the grating and the temporal nature of stimulus presentation. When presented against a uniform gray field, our results indicate that contrast sensitivity to simple wavelet subband quantization distortions displays a maximum at spatial frequencies at least as low as 1.15 c/deg. Furthermore, for the range of spatial frequencies tested here, the resulting CSF demonstrates a low-pass profile (see Fig. 6) that is consistent with what has been reported in the literature for targets of similar bandwidth. The average –3-dB cutoff frequency for CSFs surveyed by Peli et al. was approximately 3–5 c/deg for 1-octave Gabor patches (estimated by eye from Fig. 5 of Ref. 60). Our data indicate –3-dB cutoffs well within this range: 3–4 c/deg for DC and SC and 4–5 c/deg for MM.

When simple wavelet subband quantization distortions are presented against one of the two natural images used in this study, the results for subject DC indicate that the masker imposes an average threshold elevation of approximately 4–6, with minimum and maximum elevations of approximately 2–3 and 8–9 for 18.4- and 1.15-c/deg distortions, respectively. This finding is consistent with the results of Webster and Miyahara, in which detection thresholds were measured for gratings following adaptation to various natural images. Webster and Miyahara found a maximum threshold elevation (postadapt/preadapt) at 0.25–1 c/deg, a minimum threshold elevation at approximately 16 c/deg, and a threshold elevation profile not unlike those depicted in Fig. 9 (see Fig. 6 of Ref. 12). (Note that the use of natural images as maskers makes it difficult to separate the effects of masking and adaptation. Here, we have used a stimulus duration of 7 s, which, when combined with multiple trials, may very well have induced an adapted state.)

The fact that natural images are particularly effective at masking low-frequency targets is not surprising given the amplitude spectra of the images; the mean slope of the amplitude spectra of the two natural images used here was \(-1.00 \pm 0.15\) SE. This observation is illustrated in Fig. 17, in which the threshold elevation data of Subsection 4.B are replotted along with mean relative contrast energy (\(N \times c_{rms}^2\), where \(N\) denotes the size of the image) at each spatial scale computed by using the wavelet filters of Fig. 1. Note that the two trends are roughly parallel. Indeed, it would be interesting to see how well our data might be fitted by one or more of the previously described models of masking (see Subsection 2.A).

B. Summation of Visual Responses to Wavelet Distortions

Previous studies have investigated summation of responses to sine-wave and Gabor-patch targets presented against either no mask or an unnatural masker. We investigated summation of responses to wavelet distortions presented against no mask and two natural-image maskers.

In the unmasked paradigm, relative sensitivities were consistent with those found in previous summation-at-threshold studies. Summation on the orientation dimension was independent of the spatial frequencies tested (1.15, 2.3, and 4.6 c/deg), and summation on the spatial-frequency dimension was independent of the orientations tested (horizontal and vertical). Thus, for the orientations and the range of frequencies tested here, our data suggest that responses to wavelet subband quantization distortions pool in a fashion similar to that of responses to sine-wave and Gabor-patch targets, despite the fact that wavelet subband quantization distortions occupy a band of spatial frequencies as opposed to traditional, single-frequency gratings.
Relative sensitivities measured in the presence of the two natural-image maskers were greater than those found previously utilizing unnatural maskers. Bonneh and Sagi\textsuperscript{40} found RS \( \approx 1.2 \) (\( \beta \approx 4 \)) for summation across space for Gabor targets that were spatially correlated with Gabor maskers. Here, we have found RS \( \approx 1.5 \) to 1.7 (\( \beta \approx 1.3 \) to 1.6) for summation across orientation and spatial frequency for wavelet targets that were spatially correlated with natural-image maskers.

Unlike traditional masks, natural images contain relatively broadband and spatially localized frequency content. The effectiveness of a natural image at masking a particular target, thus, depends not only on the nature of the target but also on where in the image this target is located. In the context of wavelet subband quantization, because the basis functions (i.e., the wavelets) are localized both in frequency and in space, so too are the quantization distortions. Quantizing an LH subband, for example, will induce distortions that are localized to the horizontal edges within the image. Similarly, quantizing an HL subband will induce distortions that are localized to the vertical edges within the image. When both of these subbands are quantized together, the nature of the combined distortions will therefore depend on the image’s local frequency content. Consider, for example, the two segments of a natural image\textsuperscript{50} depicted in Fig. 18. Because the horizontal and vertical content of the antenna is spatially separated, so too are the components of the compound target formed when the LH and HL subbands are quantized. The image segment with the people, however, contains overlapped edges of multiple orientations; in this case, the components of the compound target are spatially superimposed. Thus, depending on where in the image observers looked, it is possible for this one natural-image masker to give rise to a range of summation values. It is therefore reasonable to assume that there exist regions within the image that are particularly effective at masking simple targets but are only marginally effective at masking compound targets, and vice versa.

One might test this latter assertion by assessing summation in the presence of a phase-randomized natural-image masker (e.g., 1/f noise). Additionally, notch-filtered image maskers might be used to test off-frequency looking, e.g., by assessing summation on the orientation dimension to components at a particular spatial scale in the presence of an image masker whose corresponding spatial scale has been filtered out. Off-frequency looking might also be tested by using compound targets containing components that are well separated on the spatial-frequency dimension such that the channel “switched to” for one component is not the same channel used for detecting the other component (e.g., by means of a compound target that consists of 1.15- and 18.4-c/deg components).

Fig. 18. Two regions of the wall image used in a previous masked-summation study\textsuperscript{50} yielding nearly linear summation. The top region, “antenna,” contains horizontal and vertical components that are spatially separated, whereas the bottom region, “people,” contains overlapped edges of multiple orientations.

Fig. 19. Balloon image reconstructed from quantized LH and HL subbands; the LL and HH subbands were not modified. In (a) the subbands were quantized such that the rms contrasts of the distortions are as specified by the thresholds of Fig. 8 for the balloon image and subject DC (i.e., \( \beta = \infty \)). In (b) the subbands were quantized such that the rms contrast of the distortions are approximately 22% of the thresholds specified in Fig. 8 (\( \beta = 1.5 \)). The distortions in (a) should be suprathreshold. These images were designed to be viewed from a distance of approximately three picture heights. Compare these images with the original balloon image depicted in Fig. 4. Pay particular attention to the sky region near the top of the balloon and to the interior of the symbol “11” located in the center of the balloon’s envelope. These images can also be viewed online.\textsuperscript{65}
C. Application to Image Compression

Figure 19 depicts the balloon image compressed66 (with the 9/7 biorthogonal filters and five decomposition levels) by using the contrast thresholds measured for subject DC and two summation rules ($\beta = \infty$ and $\beta = 1.5$). Because ten subbands (five LH and five HL) are quantized simultaneously, the resulting compound distortion contains $N = 10$ components; accordingly, we assume that $RS = N^{1/5} = 10^{1/5}$ when the distortions are proportioned such that they are equally detectable.36 Therefore, to produce a compressed image in which the compound distortion $\tilde{t}$ is at threshold, we select the contrast $C(t_i)$ of each (simple distortion) component $t_i$ as follows:

$$C(t_i) = 10^{-1/5} \times CT(t_i),$$

where $CT(t_i)$ denotes the contrast threshold of $t_i$ (measured in experiment 2; see Fig. 8).

Figure 19(a) depicts the balloon image compressed assuming a summation exponent of $\beta = \infty$. The contrasts of the distortions in this image are as specified in Fig. 8, i.e., $C(t_i) = CT(t_i) \forall i$. Figure 19(b) depicts the balloon image compressed assuming a summation exponent of $\beta = 1.5$. The contrasts of the distortions in this image are approximately 22% of those specified in Fig. 8; i.e., $C(t_i) = 10^{-0.667} \times CT(t_i) = 0.215 \times CT(t_i) \forall i$. Note that the distortions in the image compressed with $\beta = \infty$ are suprathreshold. [These images were designed to be viewed on an sRGB (Ref. 67) display from a distance of approximately three picture heights; to facilitate viewing, these and other images (compressed by using a variety of summation exponents) are available online.65]

Note that we have not tested combinations of spatial frequencies other than those described for experiments 3 and 4 (i.e., $N = 2$, $H + V$, $1.15 + 2.3$ c/deg, $2.3 + 4.6$ c/deg). It is possible that other combinations of distortions require the use of summation exponents other than those assessed here. Further testing is also required for distortions oriented at 45 and 135 deg that result from quantization of the HH subbands. In addition, the use of these results for general wavelet-based image compression requires a model that can predict thresholds on a per-image basis (see Ref. 68 for a review).

ACKNOWLEDGMENTS

The authors thank David J. Field for many useful discussions and for his help in designing the experiments. We also thank the two reviewers of an earlier draft of this paper for their helpful comments and suggestions. This work was funded by grants from the Department of Energy, the National Science Foundation, and Eastman Kodak Company; we are also grateful to Borland Software Corporation.

Corresponding author Damon Chandler can be reached by e-mail at dmc27@cornell.edu.

REFERENCES AND NOTES

25. J. M. Foley and C. C. Chen, “Pattern detection in the presence of maskers that differ in spatial phase and temporal
56. Subject MM did not participate in the parts of experiments 2 and 4 that tested summation on the spatial-frequency dimension.
65. The data and the two images used in this study are available online at http://foulard.ece.cornell.edu/dmc27/ImagEffects.html.
66. Only the LH and HL subbands have been quantized.