

Visibility of Noise in Natural Images

Stefan Winkler and Sabine Süsstrunk

Audiovisual Communications Laboratory (LCAV)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
1015 Lausanne, Switzerland

ABSTRACT

The degree of visibility of any kind of stimulus is largely determined by the background on which it is shown, a property commonly known as masking. Many psychophysical experiments have been carried out to date to understand masking of sinusoids or Gabor targets by similar maskers and by noise, and a variety of masking models have been proposed. However, these stimuli are artificial and quite simplistic compared to natural scene content. Masking models based on such experiments may not be accurate for more complex cases of masking. We investigate the visibility of noise itself as a target and use natural images as the masker. Our targets are Gaussian white noise and band-pass filtered noise of varying energy. We conducted psychophysical experiments to determine the detection threshold of these noise targets on many different types of image content and present the results here. Potential applications include image watermarking or quality assessment.

Keywords: Noise sensitivity, pattern masking, texture masking, noise masking, watermarking, quality metrics.

1. INTRODUCTION

The visibility of any kind of stimulus is largely determined by the background on which it is shown, a property commonly known as masking. In the context of compression, for example, it is helpful to think of the compression artifacts being masked by the original image. Masking explains why the artifacts are disturbing in certain regions of an image while they are hardly noticeable elsewhere. The same is true for a watermark signal inserted into an image. One of the main goals in the design and implementation of compression and watermarking systems is minimizing the visibility of these artifacts (“visually lossless compression” tries to suppress them completely) while maximizing the compression ratio or the capacity and robustness of the watermark. Therefore, it is important to understand the sensitivity of the human visual system to these artifacts under different conditions and in different masking situations.

Masking effects are usually quantified by means of detection experiments, where the contrast threshold for the detection of a target stimulus is measured for a range of masker contrasts. As masker contrasts increase, the detection threshold of the target grows as a function of the masker contrast. The exact masking behavior depends to a great extent on the types of target and masker stimuli and their relationship to each other (e.g. relative frequency and orientation). Many psychophysical experiments have been carried out to date to investigate masking of sinusoids or Gabor patches by similar maskers¹⁻⁴ and by noise,⁵⁻⁷ and a variety of masking models have been proposed.⁸ However, these stimuli are artificial and quite simplistic compared to natural scene content. Therefore, masking models based on such experiments may not be accurate for more complex cases of masking. Nonetheless, only few studies exist that investigate masking in natural images.

Eckstein et al.⁹ studied signal detection on complex backgrounds in the presence of noise for medical images and measured defect detectability as a function of signal contrast.

Watson et al.¹⁰ coined the term “entropy masking” to emphasize that the difference between contrast masking with regular structures such as sinusoidal or Gabor masks on one hand and noise masking (also referred to as texture masking) with unstructured noise-like stimuli on the other lies mainly in the familiarity of the observer with the stimuli. Natural images are somewhere in between these two extremes.

E-mail of corresponding author: stefan.winkler@epfl.ch

Kutter and Winkler¹¹ proposed a weighting mask for a spread-spectrum watermark based on masking experiments with natural images. The masking model used was still based on data from contrast masking experiments.

Nadenau et al.¹² compared masking models for the visibility of wavelet compression artifacts in natural images. Their experiments are quite specific to the application, but represent one of the first comparative studies of masking in natural images.

In this paper, we investigate the visibility of noise itself as a target on structured backgrounds such as natural images. Noise is introduced in many imaging operations — it can be due to the image sensor, quantization noise from compression, watermarking information inserted into an image, etc. We conducted psychophysical experiments to determine the detection threshold of noise targets on different types of image content. The test images and the types of noise used are presented in Section 2. The experimental setup and procedures are described in Section 3. Finally, the data we obtained are analyzed and discussed in Sections 4 and 5.

2. TEST MATERIAL

2.1. Test Images

The test images were chosen from various collections, including the JPEG2000 test set and the VisTex texture database.* We cropped square regions of 400×400 pixels from the original images and converted them to gray scale. In the selection process, we attempted to include images with various characteristics such as flat areas, complex textures, strong contours, and faces. We also added a uniform gray image to the test set to determine the noise visibility threshold in the absence of masking. The complete set of test images is shown in Figure 1.

2.2. Noise Insertion

The basis for noise insertion is additive white Gaussian noise (AWGN). To test the frequency dependence of noise visibility, we also use band-pass filtered noise. All noise patterns are shaped in the spatial domain with a Gaussian envelope to limit the image region affected by noise and at the same time to avoid any border effects of a sharp cutoff.

In the white noise case, two envelope sizes are used to test the effect of the spread of the noise over the image, one with a σ_e of 1/10 of the image width (40 pixels or 1.3 degrees of the visual field for the given viewing setup, see Section 3), and one with a σ_e of 1/5 of the image width (80 pixels or 2.6 degrees). The second filter thus retains four times the total noise energy of the first.

For band-pass filtering, we use the filters proposed by Daly for his Visible Differences Predictor.¹³ They are defined using a hierarchy of low-pass filters with the following radial frequency characteristics:

$$H_{LP}(f) = \begin{cases} 1 & \text{if } f \leq f_c - w_t/2, \\ \frac{1}{2} + \frac{1}{2} \cos\left(\pi \frac{f-f_c}{w_t} + \frac{\pi}{2}\right) & \text{if } f_c - w_t/2 < f < f_c + w_t/2, \\ 0 & \text{if } f \geq f_c + w_t/2, \end{cases}$$

where f_c is the cutoff frequency and w_t is the transition width. A band-pass filter is created by subtracting two such low-pass filters. We chose cutoff frequencies of 5 and 10 cycles per degree (cpd) for the first and 10 and 20 cpd for the second band-pass filter. The transition width was $w_t = \frac{2}{3}f_c$ in all cases. The frequency-domain representations of these two filters for the given image size and viewing setup (see Section 3) are shown in Figure 2. Note that the second filter is almost high-pass. After filtering the noise, it is again shaped with a Gaussian envelope with a σ_e of 1/10 of the image width.

The resulting test conditions are summarized in Table 1; Figure 3 shows examples for each case after insertion into a test image.

* <http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>



Figure 1. Test images, ordered from left to right and top to bottom by the visibility threshold of the unfiltered noise with $\sigma_e = 1.3^\circ$ (cf. Section 4). The top left image is uniform gray of the same luminance as the screen background in the experiments.

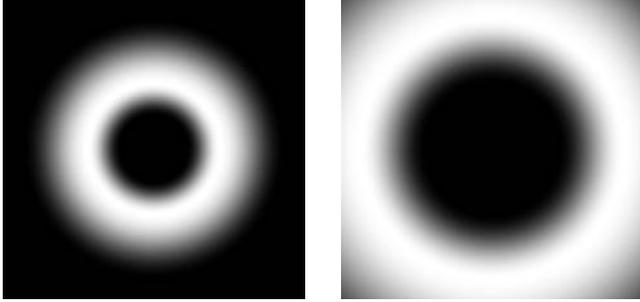


Figure 2. Band-pass filters with 5-10 cpd (left) and 10-20 cpd (right).

#	Filter	Envelope
1	none	$\sigma_e = 1.3^\circ$
2	none	$\sigma_e = 2.6^\circ$
3	5-10 cpd	$\sigma_e = 1.3^\circ$
4	10-20 cpd	$\sigma_e = 1.3^\circ$

Table 1. Noise parameters.

3. EXPERIMENTS

3.1. Presentation Structure

The presentation of the test images was based on a two alternatives forced choice (2AFC) design, i.e. subjects were shown the image with and without noise and had to choose which one contained the noise. The two images with and without noise were displayed on the left side of the screen, one in the top half, the other in the bottom half (in our display setup, we additionally showed the original image on the right side of the screen throughout the entire trial). The subjects were also given the possibility of answering “don’t know” if they could not detect the noise in either of the two images. While this works against the 2AFC idea of eliminating differences in conscious decision criteria between subjects, we introduced it to reduce the number answers needed in a trial to converge toward the threshold.

We used the PEST (Parameter Estimation and Sequential Testing)¹⁴ procedure to steer the noise level sequence for an image trial. PEST is an adaptive staircase procedure that uses maximum likelihood estimation to determine the threshold. The threshold is typically found as follows: from a given initial noise level chosen by the experimenter, the noise level is decreased by a certain step size until a wrong answer is given by the subject. When this happens, the step size is halved, and the noise level is increased again, until the right answer is given, and so on. The PEST procedure ends when the step size drops below a specified limit. In our case, a “don’t know” response was treated just like a wrong answer for the choice of the next noise level.

The images were shown on a uniform gray background. The observers did not have a time limit for answering. The maximum number of responses per trial was restricted to 40 as a precaution against cases where no convergence toward a threshold could be achieved. On average, it took subjects about 2 minutes per image to arrive at the threshold. After each response, the background was displayed without the images for approximately one second.

Two separate experiments were conducted. The first experiment contained the test images with white noise and two envelope sizes, and the second experiment contained the test images with band-pass filtered noise. In order to minimize contextual effects, the order of the test images was randomized such that every observer viewed them in a different order. The experiments were divided into separate sessions with at most 7 test images. Instructions were given to the participants in written form. After they had read the instructions, a training session was run to demonstrate the task that viewers had to perform.

3.2. Observers and Viewing Conditions

A total of 13 observers participated in the experiments, but not everyone viewed all test images. All observers had normal or corrected to normal vision. Four of them were experienced with psychophysical experiments.

The screen was a calibrated 21” Barco PCD 321+ monitor with a resolution of 1280×1024 pixels. There was no additional light source in the viewing room. During the experiments, the observers were positioned at a distance of 50 cm from the screen. This setup implies a sampling grid of 30 pixels per degree of visual angle, with an image extending over approximately 13 degrees of the field of view.



(a) White noise, envelope $\sigma_e = 1.3^\circ$



(b) White noise, envelope $\sigma_e = 2.6^\circ$



(c) Band-pass noise 5-10 cpd, envelope $\sigma_e = 1.3^\circ$



(d) Band-pass noise 10-20 cpd, envelope $\sigma_e = 1.3^\circ$

Figure 3. Test image with four different types of noise (see Section 2.2).

4. RESULTS

4.1. Subjective Data

We first examine the effects of envelope width and noise filtering on the visibility threshold of the noise. Figure 4 shows the global noise energy at threshold for all test images. The images are ordered according to the visibility threshold of the white noise with $\sigma_e = 1.3^\circ$ as in Figure 1. The general effects of widening the noise envelope or noise filtering can be seen. The wider envelope leads to an increased energy necessary for detection. This is an indication that detection is not governed by the entire image, but focuses instead on small subregions of an image. In that sense, the additional noise energy is actually “wasted” as it does not help detection.

Band-pass filtering has different effects depending on the frequency range of the filter. For the lower-frequency filter (5-10 cpd), the noise energy threshold is reduced compared to the unfiltered case. This indicates that our visual system is indeed more sensitive to this “structured” noise. It is quite consistent with contrast sensitivity measurements for sinusoids,¹⁵ which show that our visibility threshold is lowest for frequencies in that range. The other band-pass filter (10-20 cpd) has the opposite effect – it increases the noise energy threshold compared to the unfiltered case. This can again be explained with the frequency dependence of human vision as evidenced by contrast sensitivity measurements, which decreases rapidly toward higher frequencies. The white Gaussian noise lies between these two cases as it comprises all frequencies.

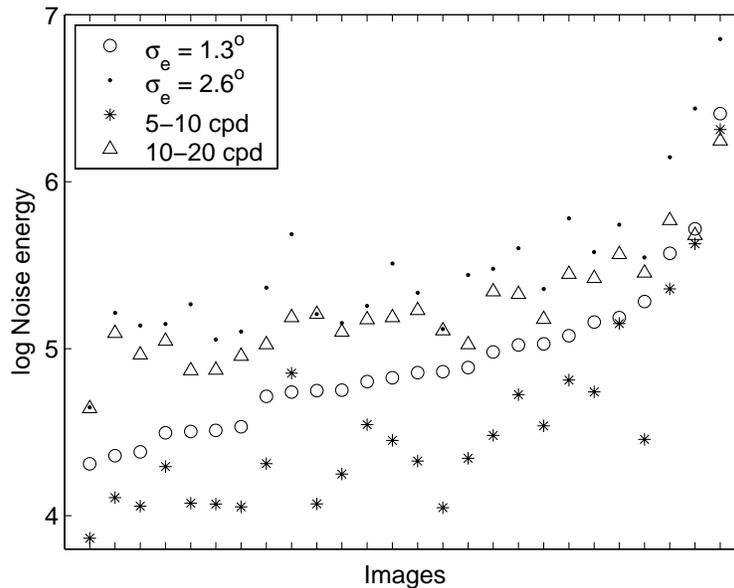


Figure 4. Noise energy at detection threshold for the different noise types and test images.

The absence of facilitation effects can also be observed. Facilitation describes a *decrease* of the detection threshold at low masker contrasts, which implies that the target can be easier to perceive in the presence of the masker than without in some cases. In our experiments, the uniform gray test image always has the lowest noise threshold; any image content only increases this threshold. As facilitation occurs mainly when target and masker have very similar properties, this could be due to noise and images not being similar enough, even though some images do appear noise-like in the region of interest. A more probable cause could be that our test set did not include images with sufficiently low activity for facilitation effects to become apparent.

4.2. Modeling

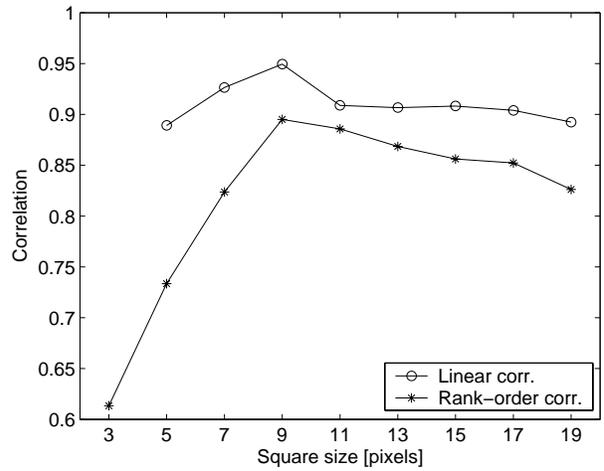
In order to predict the noise visibility for a given image, it is important to determine a well-defined functional relationship between the image and noise descriptors. Previous masking experiments yielded an approximately log-linear relationship. The difficulty with finding models for masking in natural images is the mathematical

description of the image content. We try to avoid a complex filtering approach based on human vision here and give preference to simple image analysis methods. However, we found that global image descriptors such as the overall energy are not useful for modeling our masking data.

During the experiments, subjects pointed out that they were focusing on particular spots in the image where the noise was easiest to detect. To account for this behavior, we propose the following model of masking: we compute the standard deviation of a small square region as a local measure of image activity and slide the square over the entire image to obtain a value at every pixel location. Assuming the noise visibility is highest where the image activity is lowest, we search for the square with the lowest standard deviation. To limit the search to the region with noise, it is restricted to the central part of the image defined by a circle with a radius of one tenth of the image width. This procedure is illustrated in Figure 5(a). We refer to the resulting standard deviation as $\min \sigma_{\text{image}}$. For the band-pass noise data, the image is pre-filtered with the same filter as the inserted noise prior to the calculation.



(a) Central region and optimal square size.



(b) Dependence on square size.

Figure 5. Modeling image activity. The standard deviation of a small square is computed at every pixel. Its lowest value inside the relevant image region is used as the image activity descriptor (a). Correlations between this descriptor and the threshold noise level reach their peak at a square size of around 9×9 pixels (b).

To test the validity of these assumptions, we compute the linear correlation as well as the rank-order correlation between the $\log \min \sigma_{\text{image}}$ and $\log \sigma_{\text{noise}}$ of the threshold noise level for the different experiments. Figure 5(b) shows the resulting curves as a function of the square size used for calculating σ_{image} , which gives an indication of the approximate area actually used by the observers for detecting the noise. The highest correlations are obtained for squares of 9×9 pixels and above, which corresponds quite well to the observer experience in the experiments.

The relationship between our image activity descriptor and the noise variance at threshold is shown in Figure 6 for white noise with $\sigma_e = 1.3^\circ$. As can be seen, a linear fit describes the relationship quite well; note again the absence of facilitation effects.

The parameters of the fit for all four noise types are shown in Table 2. Our model works best for white noise; the number of outliers increases for the band-pass noise cases. The slopes of the fitted lines are comparable to what was observed in masking experiments with Gabor patches and sinusoids^{1,2} as well as some of our

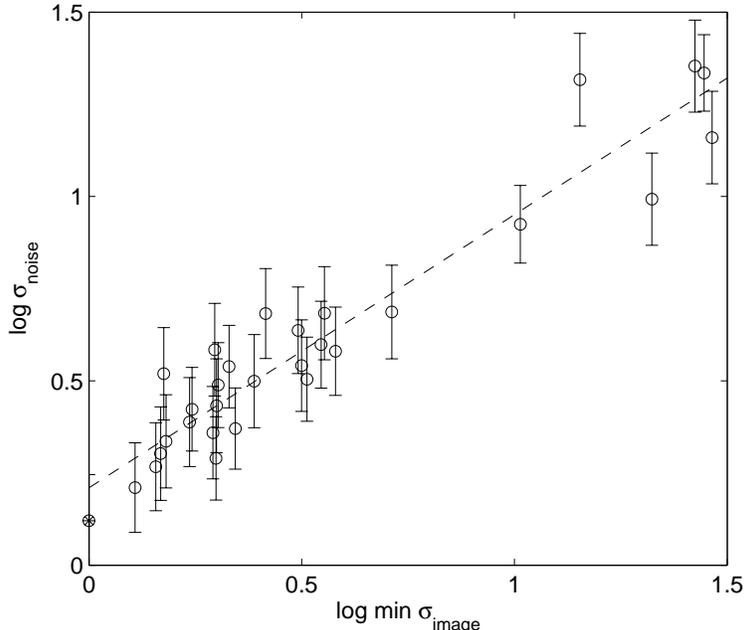


Figure 6. Masking of white Gaussian noise with $\sigma_e = 1.3^\circ$. The noise variance at threshold is shown as a function of the smallest standard deviation of a square in a central region (both on logarithmic scales). The error bars indicate the 95% confidence intervals of the observer data. The leftmost data point (denoted with a star) refers to the uniform gray image, for which $\sigma_{\text{image}} \equiv 0$; its true x-axis location in this plot would thus be at $-\infty$.

previous noise masking experiments.¹¹ The data indicate that the masking effect is least pronounced for the high-frequency band-pass noise.

Noise	Correlation	Slope
White, $\sigma_e = 1.3^\circ$	95%	0.74
White, $\sigma_e = 2.6^\circ$	93%	0.67
5-10 cpd, $\sigma_e = 1.3^\circ$	86%	0.69
10-20 cpd, $\sigma_e = 1.3^\circ$	87%	0.46

Table 2. Summary of masking model fits.

5. CONCLUSIONS

We conducted psychophysical experiments on the visibility of different types of noise in natural images. We analyzed the data and presented a simple modeling approach for activity masking. The results we obtained are consistent with data from other masking experiments in a number of ways: detection ability is higher on a plain background, and a clear masking effect was observed. In particular, noise thresholds increased significantly with image activity. Our model is able to fit the data quite well, and its parameters match previous measurements.

Nonetheless, this can only be a first step. The noise thresholds for our test images are clustered in the medium noise threshold and image activity range – we only have few data points for high masking, which can have a large influence on the slope of the fitted line. Likewise, we did not observe any facilitation effects, which may be due to a lack of test images with very little activity. The proposed masking model could be useful in selecting additional test images in these ranges. The fact that observers seem to use only a small part of the image for making their decision is reflected in the model and should be verified in experiments with smaller noise patterns and images.

Additional work is needed with respect to our model as well. As it is, it only accounts for image activity/texture masking; we have not investigated luminance masking here, which may also play an important role. Also, it incorporates knowledge about the noise characteristics (through the image pre-filtering) and only achieves individual fits to each noise type; a more general model would be preferable. In the end, the design was driven mainly by observer feedback and data analysis. It would be interesting to compare its performance with existing masking models based on the human visual system to see if better predictions could be obtained.

Finally, we have only considered luminance images and luminance noise. Studying the interactions of colored noise and natural color images will be necessary before the models can be applied in real-life applications.

ACKNOWLEDGMENTS

The work presented in this paper was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

We thank Maria Gordon for setting up and conducting the psychophysical experiments. We would also like to thank all the people who volunteered to participate in our tests.

REFERENCES

1. G. E. Legge and J. M. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America* **70**, pp. 1458–1471, December 1980.
2. J. M. Foley, "Human luminance pattern-vision mechanisms: Masking experiments require a new model," *Journal of the Optical Society of America A* **11**, pp. 1710–1719, June 1994.
3. E. Switkes, A. Bradley, and K. K. De Valois, "Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings," *Journal of the Optical Society of America A* **5**, pp. 1149–1162, July 1988.
4. C.-C. Chen, J. M. Foley, and D. H. Brainard, "Detection of chromoluminance patterns on chromoluminance pedestals. I: Threshold measurements," *Vision Research* **40**, pp. 773–788, March 2000.
5. A. van Meeteren and J. M. Valetton, "Effects of pictorial noise interfering with visual detection," *Journal of the Optical Society of America A* **5**, pp. 438–444, March 1988.
6. K. T. Blackwell, "The effect of white and filtered noise on contrast detection thresholds," *Vision Research* **38**(2), pp. 267–280, 1998.
7. K. R. Gegenfurtner and D. C. Kiper, "Contrast detection in luminance and chromatic noise," *Journal of the Optical Society of America A* **9**, pp. 1880–1888, November 1992.
8. S. A. Klein, T. Carney, L. Barghout-Stein, and C. W. Tyler, "Seven models of masking," in *Proceedings of SPIE Human Vision and Electronic Imaging*, **3016**, pp. 13–24, (San Jose, CA), February 8–14, 1997.
9. M. P. Eckstein and J. S. Whiting, "Visual signal detection in structured backgrounds. I. Effect of number of possible spatial locations and signal contrast," *Journal of the Optical Society of America A* **13**, pp. 1777–1787, September 1996.
10. A. B. Watson, R. Borthwick, and M. Taylor, "Image quality and entropy masking," in *Proceedings of SPIE Human Vision and Electronic Imaging*, **3016**, pp. 2–12, (San Jose, CA), February 8–14, 1997.
11. M. Kutter and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," *IEEE Transactions on Image Processing* **11**, pp. 16–25, January 2002.
12. M. J. Nadenau, J. Reichel, and M. Kunt, "Performance comparison of masking models based on a new psychovisual test method with natural scenery stimuli," *Signal Processing: Image Communication* **17**, pp. 807–823, November 2002.
13. S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, ed., pp. 179–206, MIT Press, 1993.
14. M. M. Taylor and C. D. Creelman, "PEST: Efficient estimates on probability functions," *Journal of the Acoustical Society of America* **41**(4), pp. 782–787, 1967.
15. F. L. van Nes and M. A. Bouman, "Spatial modulation transfer in the human eye," *Journal of the Optical Society of America* **57**, pp. 401–406, March 1967.