Perception, 2008, volume 37, pages 1098 – 1113

doi:10.1068/p5814

Uncomfortable images in art and nature

Dominic Fernandez, Arnold J Wilkins¶
Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK;
e-mail: arnold@essex.ac.uk
Received 13 September 2006, in revised form 9 November 2007

Abstract. The ratings of discomfort from a wide variety of images can be predicted from the energy at different spatial scales in the image, as measured by the Fourier amplitude spectrum of the luminance. Whereas comfortable images show the regression of Fourier amplitude against spatial frequency common in natural scenes, uncomfortable images show a regression with disproportionately greater amplitude at spatial frequencies within two octaves of 3 cycles deg^{-1}. In six studies, the amplitude in this spatial frequency range relative to that elsewhere in the spectrum explains variance in judgments of discomfort from art, from images constructed from filtered noise, and from art in which the phase or amplitude spectra have been altered. Striped patterns with spatial frequency within the above range are known to be uncomfortable and capable of provoking headaches and seizures in susceptible persons. The present findings show for the first time that, even in more complex images, the energy in this spatial-frequency range is associated with aversion. We propose a simple measurement that can predict aversion to those works of art that have reached the national media because of negative public reaction.

1 Introduction

It is known that certain geometric patterns, most notably those of stripes with a spatial frequency close to 3 cycles deg^{-1}, can induce seizures in patients with photosensitive epilepsy (Chatrian et al 1970; Wilkins 1995), and discomfort in others (Wilkins et al 1984), particularly individuals with migraine (Marcus and Soso 1989). Stripes with sine-wave luminance profile and square-wave luminance profile, and patterns comprising the first and third Fourier components of a square-wave pattern in triangular- or square-wave phase are all capable of inducing seizures when the fundamental spatial frequency is close to 3 cycles deg^{-1} (Soso et al 1980). We show for the first time that the discomfort experienced in response to more complex images can be predicted from their Fourier components, including complex images from non-representational contemporary art or created from filtered noise, and even representational images: photographs of urban and rural scenes.

Our investigation was prompted by concern over the complaints with which contemporary art is sometimes associated. In 1971 when an exhibition of Op art was held in London, the *Daily Telegraph* reported that the guards complained of headaches and were issued with dark glasses. In 1989, the front page of the *Sunday Observer* carried the story of a newspaper advert with swirling stripes that had been banned after it evoked seizures. In 2005, when a London hospital commissioned artwork and three members of staff complained of migraines in consequence, the resulting controversy reached the national press. An aversive reaction is common in response to the work of many contemporary artists (Haysom 2003; Woolls 2003), including Debbie Ayles (figure 1) who uses her migraines as an inspiration for her art (Podoll 1998).

¶ Author to whom all correspondence should be addressed.

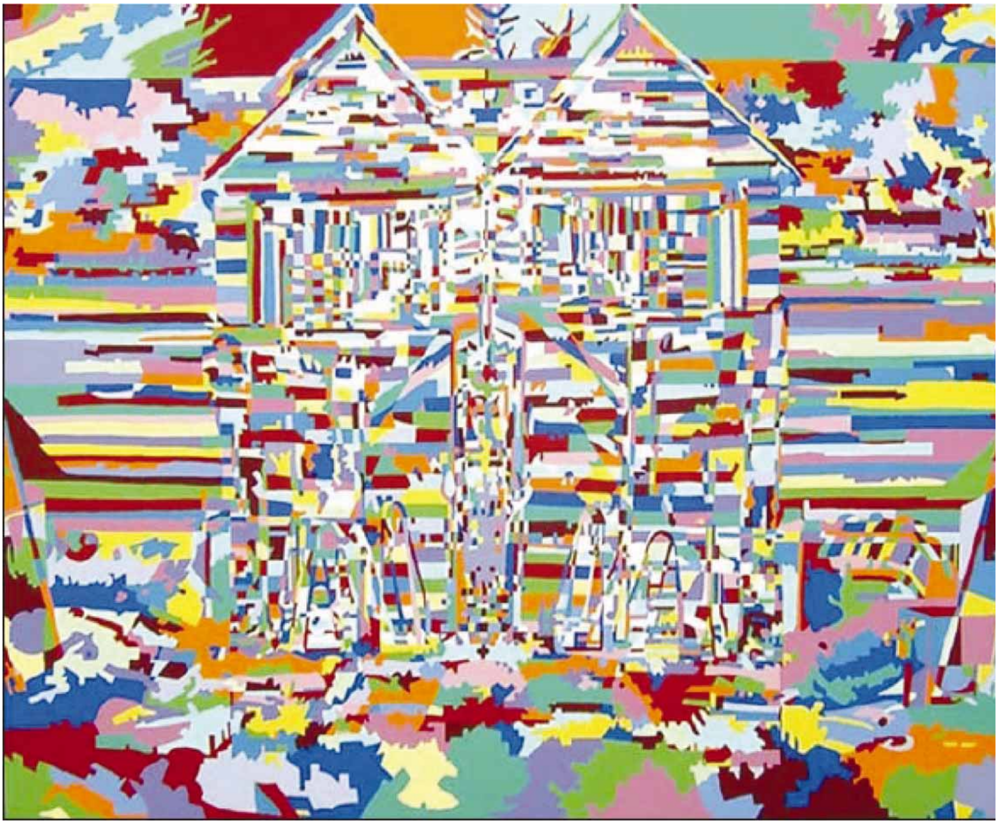


Figure 1. Jesmond Barn by Debbie Ayles© 2003 (winner of Art in Science award 2005), inspired by an attack of basilar artery migraine. Further images can be obtained from the artist's website and at http://www.migraine-aura.org/EN/Debbie_Ayles.html.

2 Preliminary studies and an introduction to methods

The work described here began with preliminary studies that were exploratory in nature. They will be described briefly, but in detail sufficient to orient the reader towards the more rigorous studies to which they gave rise.

We began by selecting nine paintings from the body of work by Debbie Ayles. The paintings were chosen as representing various genres of the artist's output. All the images used in the studies can be viewed at http://www.essex.ac.uk/psychology/overlays/sciart/. The paintings were hung as in a gallery on white walls with overhead fluorescent lighting. Three discussion groups, including artists and non-artists, observed the paintings and commented on them. We used discussion groups to minimise our preconceptions about possible sources of discomfort, and to guide hypotheses.

2.1 *Discussion groups*

The discussions were tape-recorded, the transcripts itemised and the items categorised. The following is a summary of the ideas expressed. There was a division between the artists and non-artists as to what art should do for a viewer. Artists were prepared to be visually and aesthetically stimulated, whereas non-artists wanted calm. Many participants described what they thought the paintings were representing, although the artists then criticised themselves for doing so. Most participants experienced apparent movement within the paintings which "... never let(s) the eye settle ...". Most of the images were variously described as aversive: 'intrusive', 'stressful', 'very disturbing', and 'challenging'.

The colours were generally seen as 'strident', and the "bright ones hurt (the) eyes". The experience of aversion was variously attributed to repetitive shape, to colour contrast, and to the effort in search of representational meaning. The first of these attributions was investigated in the studies that follow.

2.2 *Ratings of the artworks*

After each of the discussions, the participants were asked to observe a slide-show featuring images of the original paintings in their original coloration, and also alternative images created by the artist, similar in their spatial structure but differing in the component colours. Each of the resulting 45 images was rated with respect to both aesthetic appeal and discomfort.

2.2.1 *Participants.* Fourteen women and four men aged 28 – 66 years, mean 46 years, took part, thirteen of whom were artists. In this study the participants were tested in three small groups and steps were taken to ensure that the participants did not collude.

2.2.2 *Procedure.* Images of the artworks were presented to the three groups of participants in a constant random order as a PowerPoint slide-show. The images were projected with an Epsom EMP710 projector in an otherwise unlit room onto a 1.5 m high screen that subtended 19 – 24 deg in height, depending on viewing distance. A calibration image from the projector was measured with a Minolta TV Color Analyser II and a look-up table derived so that the on-screen luminance and CIE UCS 1976 chromaticity of each pixel could be calculated on the basis of the (uncompressed) TIFF image R, G, and B values. The slides were presented for 10 s with an interval of 6 s between each, during which a grey screen was visible with the caption "The previous slide was number ...". The participants individually rated the images on a 7-point scale arranged in columns headed 'Artistic merit' and 'Visual comfort', with left and right poles of 'Low (bad)' and 'High (good)' and 'Aversive (bad)' and 'Comfortable (good)', respectively.

3 Results

3.1 *Ratings*

The judges were consistent in rating some images as more uncomfortable than others (Kendall coefficient of concordance, $W = 0.371$, $\chi^2_{44} = 277$, $p < 0.0001$). They were also consistent (though less so) in rating some images as being of higher artistic merit than others ($W = 0.101$, $\chi^2_{44} = 79, p < 0.001$). Ratings of artistic merit correlated negatively with rated discomfort: Pearson product moment correlation coefficient, $r = -0.59$ ($p < 0.0001$) for all participants, and $r = -0.70$ ($p < 0.0001$) for artists.

3.2 *Spatial structure*

Inspection of the images rated as uncomfortable suggested that they differed with respect to spatial structure. In a preliminary analysis the spatial structure was analysed as follows. The images were re-sized, retaining the aspect ratio (using the nearest-neighbour algorithm in MATLAB®) so that the shorter dimension was 512 pixels in size; the longer dimension was then cropped so as to retain the central 512 pixels of the resized image. A Fourier analysis of the calibrated luminance of the images (512×512 pixels) was undertaken with MATLAB®, with a code incorporating that written by Kovesi (2000). The amplitude was averaged over orientation for each of the cells in the Fourier transform that had the same spatial frequency in terms of cycles per image. In this and all subsequent studies in the present report separate analyses were undertaken both with and without a Hanning filter (as defined by the MATLAB® function) to remove edge effects. The window affected the shape of the amplitude function at low frequencies but did not materially affect the function at mid-range spatial frequencies. For comparison with previous literature, we present the results of analyses without a Hanning filter.

The mean amplitude for images below and above the median comfort rating are shown as a function of spatial frequency in log–log plots in figure 2, first row, first column. The images were divided about the median rating of comfort into those that were relatively comfortable and uncomfortable. Broadly, the comfortable images showed the linear regression of log amplitude against log spatial frequency expected for natural images (Field and Brady 1997). The uncomfortable images showed a more curvilinear function with a higher amplitude at mid- to high-spatial frequencies. Figure 2, first row, second column, shows the ratio of amplitude for uncomfortable

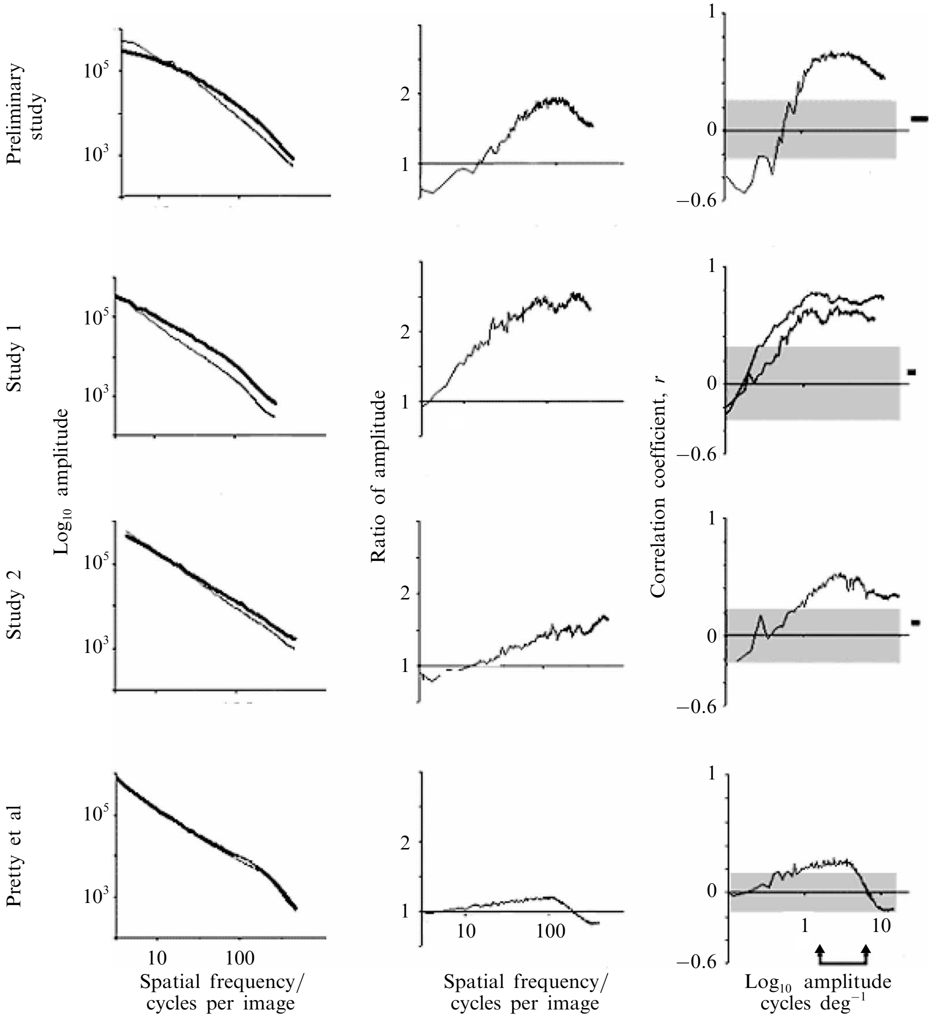


Figure 2. Rows show the results of the preliminary study, studies 1 and 2 and the study by Pretty et al (2005), respectively. First column: average $\log_{10}$ amplitude against spatial frequency for the uncomfortable images (ie those above the median rating, shown in bold lines) and comfortable images (light lines). Second column: the ratio of amplitude for uncomfortable images to that for comfortable images, shown as a function of spatial frequency. Third column: Pearson's correlation coefficient between discomfort rating and the average $\log_{10}$ amplitude at each spatial frequency. The grey shading above and below the abscissa indicates the range over which the correlation is not significant for the size of sample. The width of the horizontal bars above the abscissa indicates the tolerance in the estimate of spatial frequency, given the range of viewing distances used. The arrows indicate the range from one octave below 3 cycles $\deg^{-1}$ to one octave above.

and comfortable images as a function of spatial frequency. The difference between the comfortable and uncomfortable images was examined with greater statistical power by separately correlating over images the rating of discomfort with the average amplitude at each spatial frequency (see figure 2, first row, third column), approximating spatial frequency on the basis of the median viewing distance, and thereby introducing an error in cycles deg^{-1} no greater than 15%, shown by the horizontal bar above the abscissa. The correlation was statistically significant above 1 cycle deg^{-1} and reached a maximum of more than 0.8. (The region within which correlations are non-significant is shown shaded.)

The similarity in luminance structure of the five differently coloured versions of each of nine images may have increased the likelihood of a fortuitously high correlation. This shortcoming was corrected in the next study.

4 Study 1

In a replication of the preliminary study, 36 non-representational paintings from the Essex University collection of Latin American art and the work of the artist Cliff Colman were added to the 9 originals of Debbie Ayles, giving a total of 45 coloured images. Examples of the images are shown in the first two columns of figure 3. In this figure (but not in study 1) the images are represented in grey level and cropped so as to be square in outline.

4.1 *Methods*

4.1.1 *Participants.* Nineteen boys and twenty-eight girls, aged 15 – 16 years, from a Colchester secondary school were tested as a group.

4.1.2 *Procedure.* Presentation conditions were the same as in the preliminary study except that the images subtended 24 – 37 deg in height, depending on seating position, and the images were presented for 7 s with an interval of 7 s between each. The 7-point scales were presented individually in booklets in columns headed "How much did you like the painting?" and "Was the painting uncomfortable to look at?". The ends of the scales were headed "Not at all" on the left and "Yes, a lot" on the right. Note that the scales were, in this study, of opposite hedonic polarity (and were reversed for analysis). The participants were seated at desks, spaced to avoid collusion.

4.2 *Results*

The ratings of discomfort were concordant ($W = 0.172$, $\chi^2_{44} = 348$, $p < 0.0001$). As in the previous study, the ratings of appreciation ('liking the painting') were also concordant, though less so ($W = 0.136$, $\chi^2_{44} = 263$, $p = 0.0001$). Again, there was a significant negative correlation between the two ratings across images ($r = -0.340$, $p = 0.011$). The image analysis was similar to that in the previous study. Figure 2, second row, first column, shows amplitude for uncomfortable and comfortable images as a function of spatial frequency, and in the second column, their ratio. In the third column the spatial frequency has been calculated and, to reduce error, the participants have been divided into two groups on the basis of viewing distance, and the groups are plotted separately. The bar above the abscissa on the right of the figure shows the range of error in measurement of spatial frequency. The correlation between discomfort rating and luminance amplitude is statistically significant above 0.4 cycle deg^{-1} and at its maximum exceeds 0.7 (figure 2, first row, third column).

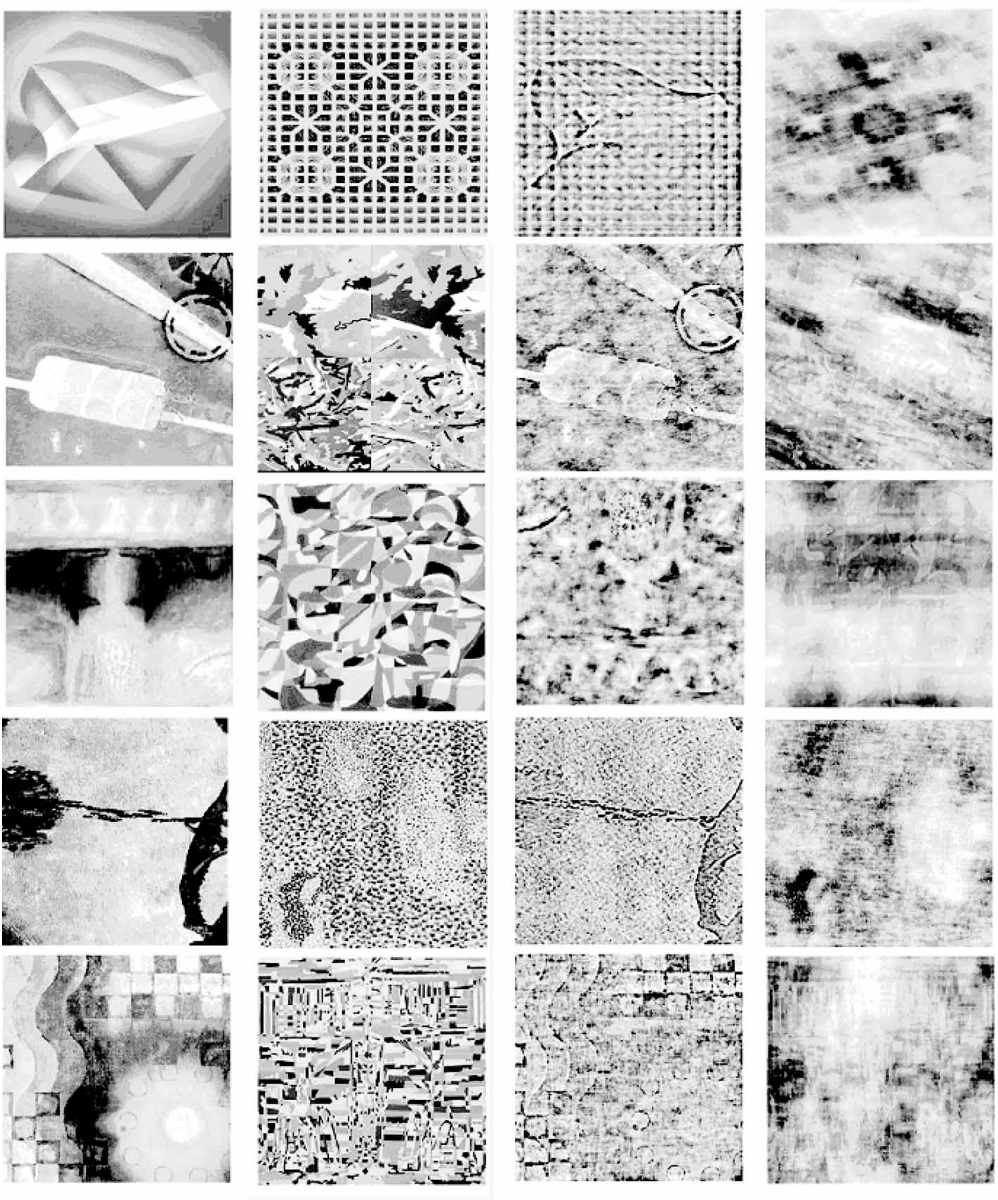


Figure 3. The first two columns show samples of the images used in study 2 in grey scale, with images rated as most comfortable in the first column and those least comfortable in the second. The third and fourth columns show images used in study 4, created by swapping the amplitude spectra of the images in the first two columns. The third column shows images having phase spectra of images that were originally rated as relatively comfortable and amplitude spectra of images that were originally rated as relatively uncomfortable. The fourth column shows images having phase spectra of the uncomfortable images and amplitude spectra of the comfortable images.

5 Study 2

In a replication of study 1, 10 non-representational paintings were chosen from the work of each of the following artists: E Gerzabek, F Casselman, K Maenz, L Taetzsch, M Brekelmans, N J Strother, and W Riggs. The images were selected at random from the published web images having sufficient size—a total of 70 images, none known to be associated with discomfort.

5.1 *Methods*

5.1.1 *Participants.* Five male and twenty-five female students from the University of Essex aged 18 – 53 years, mean 22 years, were tested as a group.

5.1.2 *Procedure.* In this study the original images were cropped so as to be 512×512 pixels in size, and care was taken to ensure that the image pixels corresponded to the projector pixels, necessitating a smaller projected image. The height of the images subtended 13 – 16 deg, depending on seating position. On this occasion the participants simply rated discomfort, and did so on a 7-point scale that varied from "Not at all" (on the left) to "Very" (on the right).

5.2 *Results*

The images included in this study were not associated with complaints of discomfort and had a smaller size than in the previous two experiments. Although the ratings of discomfort were generally lower, they were again concordant ($W = 0.212$, $\chi^2_{69} = 381$, $p < 0.0001$).

Figure 2, third row, first column, shows the amplitude for uncomfortable and comfortable images as a function of spatial frequency, and in the third row, second column, the ratio of these amplitudes. In the third column, the correlation between discomfort rating and luminance amplitude shows a similar function and is significant above 1 cycle deg^{-1}, and at its maximum exceeds 0.5. Again, the error in estimating spatial frequency is shown by the bar above the abscissa.

5.3 *Discussion*

Perhaps unsurprisingly, images rated as aversive were generally less appreciated in artistic terms. More surprisingly, studies 1 and 2 show that discomfort from complex images may be influenced by mechanisms similar to those already described in the case of simple patterns, at least in so far as the spatial-frequency tuning for discomfort is similar. The uncomfortable images have an excess of contrast energy at spatial frequencies to which the visual system is generally most sensitive. This is the case for a wide range of images and a wide range of observers and with different rating techniques. (The techniques, stimuli, and participants were deliberately changed from one experiment to the next to test the generality of observations made.)

The images used thus far were uncontrolled in terms of their spatial, chromatic, and spectral content, and, as in any correlation study, it is always possible that some unobserved covariate (for example, the number and nature of contours) was responsible for the correlation between spectral power and discomfort. The next experiments rule out this interpretation by using grey-scale images of filtered noise and manipulating the spectral power of the images displayed.

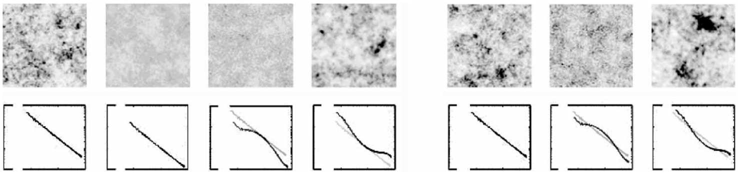
The various functions in figure 2 show the greatest association between discomfort and spectral power for spatial frequencies within two octaves of 3 cycles deg^{-1}, raising the possibility that discomfort can occur in response to complex images having a high energy at those spatial frequencies to which the visual system is generally most sensitive (ie where contrast sensitivity is greatest). If this is the case, the issue arises whether the excess of energy is best expressed in absolute terms or relative to energy elsewhere in the spectrum. The range of images over which the relationship holds suggests that the energy excess may perhaps be best expressed in relative terms, and the following studies tend to confirm this.

6 Study 3

Seven different categories of image were compared, each based upon (different) random noise. Examples of the images are shown in table 1, third row, as 'thumbnail' images.

The images in the first two categories of the seven were of noise filtered so that the logarithm of amplitude decreased linearly with the logarithm of spatial frequency, the slope of the linear function being -1, shown in the fourth row of table 1. In other words, the amplitude decreased as the reciprocal of spatial frequency, a slope that will be referred to hereafter as a 'natural slope'. Although in natural images the slope approximates -1 it must be acknowledged that the slope is seldom exactly -1 or indeed constant. In one study the slope ranged across images from -0.8 to -1.5, averaging 1.2 (Tolhurst et al 1992).

Table 1. Sample images, amplitude spectra, image statistics, mean ranks and mean ratings for the seven categories of image used in study 3. Examples of the images are shown in the third row and their spectra are shown in the fourth row. Stimuli in category N, n, and N$'$ refer to images in which the amplitude spectrum had a slope of -1. Images B and B$'$ were generated so as to have a relatively greater energy at spatial frequencies close to 3 cycles deg^{-1}. Images D and D$'$ had correspondingly lower energy at this spatial frequency. At 3 cycles deg^{-1} the energy in images B matched that in images N and the energy in images D matched that in images n (the spectra for the N and n images have been superimposed in grey on the spectra for the B and D images, respectively, to show this). The N$'$, B$'$, and D$'$ images were matched for overall energy. (Again, the spectra for the N$'$ images have been superimposed in grey on the spectra for the B$'$ and D$'$ images to show the relative magnitudes.) Rows 5–9 show various statistics of the image, including the mean and standard deviation of pixel grey level on the scale 0–1. The mean ratings have been subtracted from 10 so that high scores indicate discomfort for both ranks and ratings.



	Group 1				Group 2		
	N	n	B	D	N$'$	B$'$	D$'$
Mean	0.50	0.50	0.50	0.50	0.50	0.50	0.50
SD	0.20	0.07	0.12	0.16	0.20	0.20	0.20
Skew	0.00	0.00	0.01	-0.01	-0.04	-0.02	-0.06
Kurtosis	2.89	2.89	2.88	2.93	2.89	2.89	2.92
Total amplitude $\times 10^7$	5.12	1.72	2.95	3.43	5.18	5.02	4.41
Total power $\times 10^{11}$	1.59	1.40	1.45	1.51	1.59	1.59	1.59
Mean ranks	3.33	1.39	2.93	2.35	1.85	2.67	1.48
Mean ratings	5.54	1.54	4.63	3.50	4.98	6.46	4.04

Images in the first category, category N (N for 'natural slope'), were of high contrast and those in the second category, category n, of low.

A third category of images, category B (B for bump), was created in which the amplitude near 3 cycles deg^{-1} was increased relative to that elsewhere in the spectrum.

In the fourth category of images, category D (D for dip), the amplitude near 3 cycles deg^{-1} was decreased relative to that elsewhere in the spectrum.

The images in these categories were used to determine whether it is the absolute amplitude near 3 cycles deg^{-1}, or the unnatural curvature of the amplitude spectrum that best predicts viewing discomfort. Table 1, third row, shows examples of the images (labelled group 1) together with graphs beneath (fourth row) that relate their amplitude to spatial frequency on log–log axes.

Three further categories of image were created (labelled as group 2 in table 1). These were generated as for categories N, B, and D, but then constrained to have the same mean luminance and standard deviation of pixel grey level, and hence the same total contrast power. As a result, the difference between the high- and low-contrast 'natural slope' images (categories N and n in group 1) was removed, leaving three categories of images with spectra shown in table 1, third row: category N$'$, category B$'$, and category D$'$ (collectively, group 2). Matching the basic pixel statistics of the images meant that the effect on viewing discomfort of the distribution of contrast amplitude across spatial frequencies could be examined independently of total contrast energy.

6.1 *Methods*

6.1.1 *Stimuli.* The images were created by generating different 1024 × 1024 pixel white-noise grey-scale images that were filtered in the Fourier domain before an inverse FFT was used to create the base images with the desired amplitude spectra. The category N images were created with a radial filter which ensured that the amplitude of the noise varied with the reciprocal of the spatial frequency at all orientations. This was then normalised to give a mean of 0.5 and standard deviation of 0.2 on a linear scale of pixel values ranging from 0 to 1. This normalised image was used to create all the other categories of image.

The category n images were created by transforming the images in the Fourier domain, as above, and decreasing the amplitude by a factor of three at all spatial frequencies and all orientations.

The category B images were created by transforming category N images in the Fourier domain, filtering with a factor that conformed to a radial raised cosine with a peak centred on 3 cycles deg^{-1}. This was achieved by multiplication in the Fourier domain with a radial raised cosine (symmetric in log axes) with a peak at 76.4 cycles per image and a width of three octaves either side. The amplitude at 3 cycles deg^{-1} was the same as that in the high-contrast images (category N), but at all other spatial frequencies the amplitude was lower and, in consequence, the total contrast, standard deviation, amplitude, and power were also lower. [The cosine filter was only applied to those spatial frequencies that could be represented at all orientations (eg up to 512 cycles per image) to prevent any anisotropies in the 2-D amplitude spectrum, and this necessitated cropping of the function at the upper extreme where its value was less than 2% of the maximum value.]

The category D images were created by transforming the category N images in the Fourier domain, decreasing the amplitude in the spectrum by the same proportion in which the amplitude was increased in the category B images.

The last three categories of image, those in group 2 in table 1, were created in the same way but, following the inverse FFT, the pixel grey levels were normalised to have a mean of 0.5 and standard deviation of 0.2 on a linear scale in the range 0 – 1. Table 1 shows the statistics of pixel grey levels of the seven classes of image. Note that the skew and kurtosis are closely matched.

A Hewlett Packard LaserJet 4100 printer was calibrated with a Monolite system for surface reflection measurement with 45°/90° geometry. The experimental images (0.18 m × 0.18 m) were generated in MATLAB®, transformed with a polynomial that took account of the *K* correction, and a further polynomial based upon the printer calibration that took account of the printer nonlinearities. The calibration ensured that the pixel value of 0.5 (0 – 1 scale) had a reflectance midway between that of the black ink and the white paper and that the standard deviation of pixel values was 0.2. The transformed images were saved as TIFF files and printed from Adobe Photoshop, 0.18 m × 0.18 m in size, in a single print run.

6.1.2 *Participants.* Nine male and fourteen female Essex University students aged 18–29 years, mean 20 years, were sampled opportunistically, examined individually, and were offered no inducements to participate.

6.1.3 *Procedure.* Multiple versions of the seven categories of image were generated, each from different noise. The images were presented in groups that comprised 4 non-normalised images (group 1) or 3 normalised images (group 2) as shown in table 1. Half the participants viewed one version of the normalised images followed by the non-normalised images and half experienced the reverse order. To check for consistency, different versions of the images were then shown in the same sequence of group.

Participants were asked to hold each image at arm's length at which distance the images subtended about 25 deg. They viewed the image for 5 s and judged how comfortable it was. After viewing all the images in a group as many times as they wished they were asked to place them in a pile in order of comfort.

Having ranked all the images in each group in this way, the participants were asked to view the images again in the same sequence of groups but in a random order within each group, and to give each image a mark out of 10, where 10 was most comfortable and 1 the least. It was emphasised that their previous responses were not relevant, that this was not a test of memory, and that each image should be considered in isolation and on its own merit.

6.2 *Results*

6.2.1 *Ranks.* Each group of images, group 1 and group 2, was presented twice with different versions of the images. Participants were highly consistent in their rankings in each presentation of the images. For the images in group 1, eight of the twenty-three participants gave the same rankings for each image version, and in ten the ranks differed in only one respect. For the images in group 2, twelve participants gave identical ranks and only in two did the rankings differ completely.

Separate Friedman two-way analyses of variance by ranks revealed highly significant differences between the images in each group. The data from the two presentations were combined and are shown in table 1. Pairwise comparisons of the mean ranks of the images in group 1 by Wilcoxon tests with Bonferroni correction revealed that (i) the category N images were less comfortable than the remaining three, and (ii) category N images did not differ from the category B images whereas (iii) the category D images did. There was no significant difference between the category B and category D images (the latter had higher overall contrast). Similar separate Wilcoxon tests with Bonferroni correction comparing the 3 pairwise combinations of images in group 2, in which overall contrast was controlled, showed that the category B' images were significantly less comfortable than the other two categories of images (N' and D'), and that there was no significant difference between these.

6.2.2 *Ratings.* The image comfort ratings were consistent with the rankings and are shown in table 1. Univariate analysis of variance of the ratings in group 1, with order and image type as fixed factors, revealed no significant effect of order, but a significant difference between image categories ($F_{3,176} = 44.1$, $p < 0.0001$), which accounted for 5.1% of the variance. A posteriori Scheffé comparisons showed no difference between images in categories N and B, but a difference between those in categories N and D, which differed from one another. A similar analysis of the ratings in group 2 again revealed no significant effect of order, but a significant difference between image categories ($F_{2,132} = 20.7$, $p < 0.0001$), which accounted for 3.6% of the variance. A posteriori Scheffé comparisons revealed significant differences between all image categories, indicating that images in category B' were more uncomfortable than those in category N', which in turn were more uncomfortable than those in category D'.

6.3 *Discussion*

Images with high amplitude at 3 cycles $\deg^{-1}$ were more uncomfortable than those with similar standard deviation and hence total contrast power throughout the spectrum, and these in turn were more uncomfortable than images in which the energy at 3 cycles $\deg^{-1}$ had been reduced. Noise images with high contrast (category N) were consistently ranked and rated as less comfortable than those with low contrast (category n). Images with lower contrast overall but similar amplitude at 3 cycles $\deg^{-1}$ were as uncomfortable as those with higher contrast. These findings demonstrate that it is the distribution of energy throughout the spectrum that is critical in determining discomfort, and in particular the energy at about 3 cycles $\deg^{-1}$ relative to that elsewhere.

7 Study 4

In this study we tested the generality of the above findings with grey-scale images from the contemporary art used in study 2. The phase and amplitude spectra of the 5 most comfortable and 5 least comfortable images in study 2 were exchanged to create 10 new images. The original images and their transformations are shown in figure 3.

7.1 *Methods*

7.1.1 *Stimuli.* The five most comfortable and five least comfortable images from study 2, shown in figure 3 (first and second columns) were cropped symmetrically to give a square aspect ratio, 512×512 pixels in size, and converted to grey-scale (luminance) values by means of the MATLAB® rgb2gray function. The amplitude and phase spectra were then exchanged between a comfortable and an uncomfortable pair of images. The pairs were selected on the basis of the mean ratings in study 2, as follows: the most uncomfortable image was paired with the image given the fifth most comfortable rating; the second most uncomfortable image was paired with the fourth most comfortable, and so forth, so as to maintain a rank difference of 40 in discomfort ratings of images whose amplitude spectra were exchanged.

The above transformation resulted in 10 'chimeric' images, shown in the third and fourth columns of figure 3. The third column shows images created from phase spectra of images previously rated as comfortable, and amplitude spectra of images previously rated as uncomfortable. All images were normalised to have a mean of 0.5 and a standard deviation of 0.2 on a grey scale ranging from 0 to 1. On this scale, the skew and kurtosis of the images in the third column were 0.06 (SD 0.14) and 3.92 (SD 0.41), respectively; and those in the fourth column, 0.17 (SD 0.10) and 2.76 (SD 0.50), respectively. There were no significant differences in skew, but a significantly larger kurtosis for the images in the third column.

Images were then printed with the use of the same calibration transformations, printer specifications, and image size as in the previous study.

7.1.2 *Participants.* Twenty six female and twenty male school pupils aged 14 – 15 years participated as a group, as part of a science class.

7.1.3 *Procedure.* The 10 images were presented one on each page of a booklet in an order separately randomised for each participant, to prevent collusion. As in the previous study, participants were asked to hold the images at arm's length for 5 s and to rate the images for viewing comfort (10/10 as most comfortable, 1/10 as least) by writing the appropriate number below the image. Again it was stressed there were no right or wrong answers and that individual subjective judgment was important.

7.2 *Results*

The mean rating of the images with phase spectra of images originally rated in study 2 as relatively comfortable (figure 3, third column) were scored as more uncomfortable, mean rating 5.36 (SD 2.64). The mean ratings of the images with phase spectra of those originally rated as uncomfortable (figure 3, fourth column) were scored as relatively comfortable, mean rating 6.15 (SD 2.57). A repeated-measures ANOVA with image group and image pair (1 – 5) as factors revealed a significant main effect of group ($F_{1,41} = 7.26$, $p = 0.01$). None of the other terms approached significance. An analysis across images showed a significant difference between image group ($t_4 = 2.77$, $p < 0.05$). The results indicate that it is the amplitude spectra of the images and not their phase spectra that influence ratings of discomfort.

7.3 *Discussion*

The results of study 4 suggest that the amplitude spectrum of an image influences its discomfort more than the phase spectrum, even though the structural appearance of the image is relatively preserved in the phase spectrum. The preservation of the phase spectra resulted in (grey-scale) images that resembled the originals, although the original images as such were not rated.

The final study included the original artwork rendered as grey-scale images, along with versions in which the amplitude spectra had been transformed so as to have 'natural slope', or spectra with amplitude increased at about 3 cycles deg^{-1}.

8 Study 5

Three versions of the images shown in the first two columns of figure 3 were compared. All had a mean of 0.5 and a standard deviation of 0.25 on a grey scale of 0 – 1. One version was the original work of art in grey-scale form. On the 0 – 1 scale, the average skew of these images was 0.08 (SD 0.66), and average kurtosis 2.55 (SD 0.94). The second versions of the images were generated so as to have a 'natural slope'— average skew -0.70 (SD 0.39), average kurtosis 4.58 (SD 1.96); in the third version, the amplitude spectrum was increased near 3 cycles deg^{-1}—average skew -0.79 (SD 0.29), average kurtosis 6.64 (SD 3.33). The skew of the second and third versions did not differ significantly. All versions differed with respect to kurtosis.

8.1 *Methods*

8.1.1 *Stimuli.* A noise image was generated and filtered to have an amplitude spectrum with 'natural slope', as described in study 3. The amplitude spectrum of this noise image was exchanged with that of the original image, as described in the previous experiment. This gave an image with the same phase as the original but an amplitude spectrum that varied as the reciprocal of spatial frequency.

Using the above method, for each original artwork, the spectrum of the 512 × 512 pixel grey-scale image was exchanged with the spectrum of a 512 × 512 pixel noise image in which the amplitude spectrum had been increased by a factor of 3 at 76.4 cycles per image, about 3 cycles deg^{-1}, as described for category B images in study 3 above.

All images were normalised to have a mean of 0.5 and a standard deviation of 0.25 on a grey scale ranging from 0 to 1 and were produced with the same specifications as in the previous two studies.

8.1.2 *Participants.* Twenty eight females and five males, aged 17 – 59 years, mean 21.7 years, took part during a visit to the University of Essex. They participated as a group.

8.1.3 *Procedure.* Participants were presented with ten separate three-page booklets containing the three versions of each image. They were required to view the pages as previously described, and rank them from 1 to 3 in decreasing order of viewing comfort. They did so for all ten booklets, presented in random order.

8.2 *Results*

The judges were consistent in ranking some images as more comfortable than others (Kendall coefficient of concordance, $W = 0.375$, $\chi^2_{29} = 359$, $p < 0.0001$). The mean rank of the original versions was 1.56, that of the versions with 'natural slope' 1.89, and those with increased amplitude at 3 cycles deg^{-1} 2.57, on average. Analysis of variance of ranks averaged over participants showed a highly significant effect of image type across images ($F_{2,18} = 20.9$, $p < 0.001$). Analysis of variance of ranks averaged over images showed a highly significant effect of image type across subjects ($F_{2,64} = 38.2$, $p < 0.001$). Individual pairwise *t*-tests with Bonferroni correction showed that, across participants, all image types differed significantly. This was also the case across images, with the exception of the original and the images having amplitude spectra with 'natural slope'.

Note that, although the images with high kurtosis were ranked as less comfortable in this experiment, this was not the case in the previous experiment, even though in both experiments the images with greater energy at 3 cycles deg^{-1} were classified as least comfortable.

9 General discussion

Simple patterns of stripes with a spatial frequency within two octaves of 3 cycles deg^{-1} can, in those who are susceptible, induce seizures (Wilkins et al 1979) and headaches (Huang et al 2003; Marcus and Soso 1989). Similar patterns induce discomfort in normal individuals (Wilkins et al 1984), discomfort to which individuals with migraine are particularly susceptible (Marcus and Soso 1989). (The susceptibility is consistent with other convergent evidence for a cortical hyperexcitability in migraine—Welch 2003.)

In all six studies the amplitude of the Fourier spectrum at spatial frequencies within an octave of 3 cycles deg^{-1} (the range shown between the arrows in figure 2, third column) correlated with ratings of aversion, and the similarity between studies is evident from the figure. Although in studies 4 and 5 the increased amplitude was associated with an increased kurtosis, this was not the case in study 3, where the kurtosis was similar for all images.

Figure 4 shows the correlation between discomfort and amplitude, averaged across the curves shown in the third column of figure 2, together with functions that show the way in which the spatial frequency of a square-wave grating affects the probability of seizures in patients with photosensitive epilepsy, and aversion in normal observers.

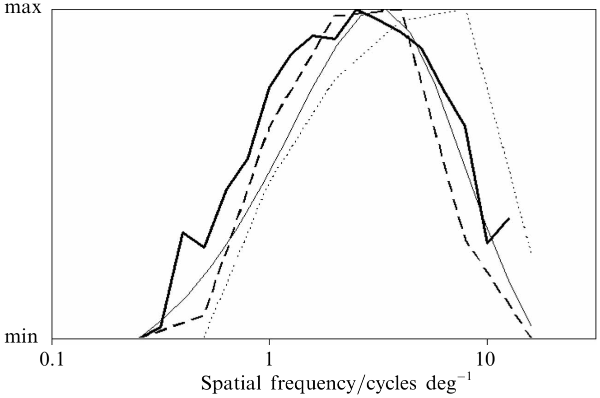


Figure 4. The correlation between discomfort and $\log_{10}$ amplitude, averaged across the curves shown in figure 2, third column (bold continuous line), together with functions that show the way in which the spatial frequency of a square-wave grating affects the probability of seizures (broken line) in patients with photosensitive epilepsy (after Wilkins et al 1979), and aversion (dotted line) in normal observers (after Wilkins et al 1984). The contrast sensitivity function for sine-wave gratings (light continuous line) is also shown (after Barten 1999). The functions have been adjusted to have the same values of maximum and minimum within the range of spatial frequency shown.

Figure 4 also shows the contrast sensitivity function for sine-wave gratings subtending 20 deg with a mean luminance that of the average across studies, 24 cd m^{-2}, according to the model of Barten (1999). The maxima are similar, as are the low-frequency arms of the functions, despite the mixture of luminance profiles (square-wave in the case of seizures and aversion, sine-wave in the case of the contrast sensitivity function, and complex in the case of the correlation). The high-frequency arms of the functions are similar, which makes it unlikely that the present findings can be interpreted simply in terms of the number of contours in the images. The number of contours in a high-frequency grating increases with spatial frequency. The similarity of the findings across studies would tend to confirm that it is the spatial frequency expressed in cycles deg^{-1} rather than cycles per image that is critical in determining aversion.

9.1 *An index of aversion?*

The correlation function in figure 4 peaks at a spatial frequency close to 3 cycles deg^{-1}. Gratings with a similar spatial frequency are those most likely to induce seizures and discomfort, whereas gratings with a spatial frequency three octaves below do not cause any seizures or discomfort (Wilkins et al 1979, 1984). We therefore calculated the ratio of the log amplitude near 3 cycles deg^{-1} to that three octaves below, to obtain an index that was relatively unaffected by calibration. (We averaged the amplitude for spatial frequencies in bins ranging over a factor equal to $\sqrt{2}$ and chose the bins nearest 3 and 0.375 cycles deg^{-1}.) The ratio averaged 0.69 (SD 0.07), 0.64 (SD 0.11), 0.73 (SD 0.09), and 0.73 (SD 0.04), respectively, in the preliminary study, studies 1 and 2, and an additional study by Pretty et al (2005), described below. The correlation between this index and ratings of discomfort was 0.72 ($p < 0.0001$, non-directional), 0.54 ($p = 0.0001$), 0.26 ($p = 0.030$), and 0.22 ($p = 0.016$), respectively (see figure 5

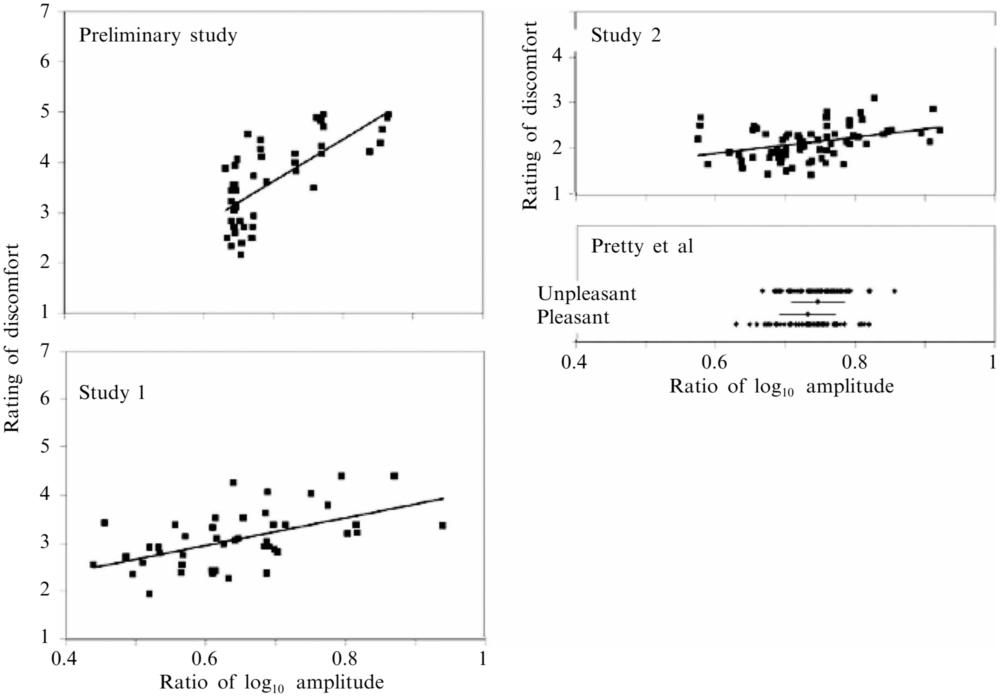


Figure 5. Scatterplots (including the line of least-squares fit) showing the correlations obtained in four studies between the mean rating of discomfort of an image, and the ratio of $\log_{10}$ amplitude at 3 cycles deg^{-1} relative to that at 0.375 cycles deg^{-1}. The data for the study by Pretty et al (2005), in which the images were classified as comfortable and uncomfortable, are shown not only by points, but also by bars indicating means ±1 SD, the upper line for the uncomfortable images.

for scatterplots). The correlation was greatest in the preliminary study and study 1, which included art that had been associated with complaints. However a significant proportion of the variance was also explained in the case of other art (study 2). We calculated the ratios for the images that caused the press attention referred to in the introduction. The ratios lay in the range 0.86 – 1.05, all more than two standard deviations above the mean for the images in the above studies.

9.2 *Generalisation to environmental images?*

For a study of exercise physiology, Pretty et al (2005) selected (from more than 300 photographs) a set of rural and urban scenes that were consistently rated as either pleasant or unpleasant: 30 photographs for each combination of pleasant/unpleasant and urban/rural, subtense 17 deg. The authors kindly made these 120 images available for analysis. The images were analysed in the same way as those in the first two studies. As can be seen from figure 2, last row, third column, the correlations (in this case point-biserial correlations) between unpleasantness and difference in amplitude were significant in the range 0.8 – 7 cycles deg^{-1}, and at maximum exceeded 0.25. The aversion index described in the previous section was significantly higher for the unpleasant than for the pleasant images (see figure 5).

10 Conclusion

The images we have studied vary in the extent to which their statistical properties resemble those of natural images, and, conversely, in the extent to which they are regarded as uncomfortable. In a wide range of images it has proved possible to predict ratings of discomfort by using the amplitude of the Fourier spectrum at a spatial frequency within two octaves of 3 cycles deg^{-1}. The present findings are consistent with the idea that the visual system has evolved to process natural scenes efficiently and that images with unnatural statistics can sometimes be stressful physiologically, particularly when they have an excess of energy at spatial frequencies to which the visual system is generally most sensitive. We suggest that the statistics of uncomfortable images revealed here could be used to avoid the controversies that result when stressful images appear in contexts where such images are inappropriate, as, for example, in public art, particularly art in hospitals.

Acknowledgments. We thank the artists who contributed their images, particularly Debbie Ayles. We thank the students and staff at St Helena School, Colchester, UK for participating, Peter Bex and Adrian Clark for guidance with the programming, and Horace B Barlow, Peter A Lawrence, Simon Laughlin, and anonymous referees for comments on an earlier draft. The preliminary work was supported by a Wellcome Trust Sciart Research and Development award to Debbie Ayles and Arnold Wilkins.

References

Barten P G J, 1999 *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality* (Knegsel: HV Press)

Chatrian G E, Lettich E, Miller L H, Green J R, 1970 "Pattern-sensitive epilepsy. I. An electrographic study of its mechanisms" *Epilepsia* **11** 125 – 149

Field D J, Brady N, 1997 "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes" *Vision Research* **37** 3367 – 3383

Haysom R, 2003 "Abstract art: pain and discomfort" *Double Dialogues* issue 4, http://www.doubledialogues.com/archive/issue_four/haysom.htm

Huang J, Cooper T G, Satana D, Kaufman D I, Cao Y, 2003 "Visual distortion provoked by a stimulus in migraine associated with hyperneuronal activity" *Headache* **43** 664 – 671

Kovesi P, 2000 *MATLAB and Octave Functions for Computer Vision and Image Processing* School of Computer Science & Software Engineering, University of Western Australia: http://www.csse.uwa.edu.au/~pk/research/matlabfns

Marcus D A, Soso M J, 1989 "Migraine and stripe-induced visual discomfort" *Archives of Neurology* **46** 1129 – 1132

Podoll K, 1998 "Migraine art—the migraine experience from within" *Cephalalgia* **18** 376

Pretty J, Peacock J, Sellens M, Griffin M, 2005 "The mental and physical health outcomes of green exercise" *International Journal of Environmental Health Research* **15** 319 – 337

Soso M J, Lettich E, Belgum J H, 1980 "Pattern-sensitive epilepsy. I: A demonstration of a spatial frequency selective epileptic response to gratings" *Epilepsia* **21** 301 – 312

Tolhurst D J, Tadmor Y, Chao T, 1992 "Amplitude spectra of natural images" *Ophthalmic and Physiological Optics* **12** 229 – 232

Welch K M, 2003 "Contemporary concepts of migraine pathogenesis" *Neurology* **61** S2 – S8

Woolls D, 2003 *Abstract Art Used to Drive Prisoners Mad* (Associated Press) http://www.iol.co.za/index.php?set id=1&click id=3&art id=qw1048861 440340B215

Wilkins A J, 1995 *Visual Stress* (Oxford: Oxford University Press)

Wilkins A J, Darby C E, Binnie C D, 1979 "Neurophysiological aspects of pattern-sensitive epilepsy" *Brain* **102** 1 – 25

Wilkins A J, Nimmo-Smith M I, Tait A, McManus C, Della Sala S, Tilley A, Arnold K, Barrie M, Scott S, 1984 "A neurological basis for visual discomfort" *Brain* **107** 989 – 1017