Less is more: Subjective detailedness depends on stimulus size

Sven P. Heinrich

Sektion Funktionelle Sehforschung, Univ.-Augenklinik
Freiburg, Freiburg, Germany

Michael Bach

Sektion Funktionelle Sehforschung, Univ.-Augenklinik
Freiburg, Freiburg, Germany

Subjective detailedness is the spontaneously perceived overall detailedness of an image. Cursory experience suggests that, for identical objective detailedness, subjective detailedness is higher in small than in large images. While previous studies hint towards a size-dependence of subjective detailedness, they are not conclusive regarding the magnitude, direction and robustness of the effect, in particular in natural images. Subjects performed a two-interval forced choice task, deciding which of a pair of images they perceived as more detailed. We used both natural images and random patterns. One of the images in each pair was scaled down to one third of the linear size of the other, but was otherwise identical. Objective detailedness was adjusted for each image individually via a low-pass filter and the point of subjective equivalence was determined. Depending on the type of image, subjects required the objective detailedness of the larger image to be 0.26–0.48 log units (median values) higher than that of the smaller image for perceiving the same subjective detailedness. This implies that, with equal objective detailedness, smaller images appear more detailed. Our results demonstrate that the effect is sizable and robust across most subjects and across repetitions. It is not readily explained by differences in contrast sensitivity, but may rather have its origin in visuo-cognitive strategies of image evaluation.

Keywords: detailedness, natural images, spatial sensitivity, but may rather have its origin in visuo-cognitive strategies of image evaluation.

Introduction

Imagine watching a severely out-of-focus printed photograph, holding it at arm’s length with image blur greatly reducing the details that can be recognized. If you now tack the photograph to the wall and step back to watch it from a moderate distance, say a couple of meters, the blur due to defocus may still be the limiting factor for recognizing details. Nevertheless, many people would agree that the blur is less disturbing now, and the picture seems to be more detailed than before. In the age of digital photography, it is easily possible to observe the effect on a computer screen by resizing the image, which shows that the effect does not depend on varying the distance at which a photograph is seen, but rather on the visual angle spanned. In the present study we aimed at confirming and quantifying these cursory findings under laboratory conditions.

Subsequently, we will use the terms “subjective detailedness” and “objective detailedness” to denote the subject’s perception and the physical stimulus properties, respectively. A rating of subjective detailedness is not obtained by counting individual stimulus elements, but rather reflects a spontaneous overall impression. Objective detailedness in natural images can be controlled by a spatial low-pass filter.

The phenomenon investigated here is not the only case where a subjective increase in image quality is perceived without any information added to the image. A well-known example is the mosaic effect (Harmon & Julesz, 1973), where pictures are assembled from large homogenous blocks. There, blurring the image (or watching it from a far enough distance) removes the influence of the block boundaries and facilitates the recognition of the picture content. As opposed to these effects, the phenomenon investigated in the present study only depends on the image scaling, but not on the change of any image content.

Subjective detailedness may be seen as a facet of subjective image quality. The latter is frequently assessed in photography and imaging science and can be measured subjectively or objectively (Nilsson, 1999). It has been proposed to describe image quality on three levels, namely image quality metrics, image quality attributes, and image preference (e.g., Dalal, Rasmussen, Nakaya, Crean, & Sato, 1998; Natale-Hoffman, Dalal, Rasmussen, & Sato, 1999). Image quality metrics are primarily simple physical measures (such as line density), though they may be scaled to account for human perception. Image preference is a subjective rating of like or dislike, usually on a one-dimensional scale. Image quality attributes describe an intermediate level of image quality, such as overall line quality, combining both subjective and objective characteristics.

Turning from imaging science to vision science, one may wish to devise a visuo-cognitive scheme of image quality measures. In addition to a physical level of description, an
intermediate level would account for the properties of the human visual system while being largely agnostic to the meaning of the image content. Subjective detailedness, together with other measures, would be a candidate for such an intermediate descriptor. The top level of image quality measures would describe how well objects or other meaningful image features can be detected or recognized. The relationship between the lowest and the highest level of description is subject to many investigations, for instance in the field of radiology (van Overveld, 1995). Ultimately, one may envisage rules that transform one level of description into a different level through mathematical coordinate transformation. Such a conversion has already been successfully demonstrated in some specific cases (e.g., Martens, 2002; Martens & Kayargadde, 1996).

A question germane to the present one was investigated by Barten (1989), who assessed subjective ratings of image quality and found that there is an optimal display size (or optimal viewing distance) for a given number of pixels. The optimal conditions can be computed with the aid of an equation that takes into account the modulation transfer functions of the eye and of the display device. The subjective rating of image quality is inherently difficult, not only because it usually involves the attempt to map a multifactorial problem onto a one-dimensional scale, but also because there is a sizable interindividual variability. Even with blur alone, subjects differ considerably regarding their threshold for perceiving the image degradation (Layton, Dickinson, & Pluznick, 1978).

A number of further studies have investigated aspects of vision that are related to the present question. Field and Brady (1997) and Tolhurst and Tadmor (1997), among others, have assessed the perception of blur in different contexts, aiming at a better understanding of the mechanisms underlying perception, and Schieber (1994) measured the blur tolerance to optimize the legibility of highway signs. These studies, however, do not predict subjective detailedness. Similarly, previous studies by, for instance, Nothdurft (1985), Joseph, Victor, and Optican (1997), Kingdom and Keeble (1999), and Rainville and Kingdom (2002) reveal important characteristics of the scale invariance of specific perceptual tasks, such as texture segregation and the perception of spatial correlation, without providing a basis for the prediction of subjective detailedness. Interestingly, though, Parish and Sperling (1991) did not find an effect of viewing distance, i.e., retinal size, on the discrimination of letters that were spatially filtered and had noise added.

In a thorough series of investigations, Vicario (1971a, 1971b, 1972) and Masin (1980) found that a grating covering a small stimulus area can be perceived clearly and in full at a shorter distance than a grating of identical spatial frequency covering a larger stimulus area, and that the density of the grating appears lower in the small stimulus. This finding predicts that a small stimulus should appear less detailed than a large stimulus. Another related area of visual performance is the perception of numerosity, which has recently gained renewed interest, and which has been proposed as an independent primary visual property (Burr & Ross, 2008, but see also Durgin, 2008). Already Ponzo (1928) demonstrated with a series of visual illusions that perceived numerosity may depend on a variety of geometrical factors including the size of the stimulus. For instance, he showed that a small spoke wheel appears to have more spokes than a large spoke wheel, despite the number of spokes objectively being the same. Although in this case the direction of the effect indeed seems to support our prediction of a higher subjective detailedness in small stimuli, the results of experiments on numerosity cannot necessarily be transferred to the question of detailedness of natural stimuli as those usually contain a small number of main elements (a few people, some houses, or similar) which would not be perceived as changing in number even when detailedness changes.

In summary, the idea that subjective detailedness differs from objective detailedness is hinted at in previous work, but it needs to be assessed whether it is a substantial and robust phenomenon. This, and a first explanatory attempt, is the aim of the present report.

Methods

We performed two experiments that shared most methodological details. In the first experiment, we demonstrated the effect with various natural images and with random patterns. In the second experiment we investigated how the effect depends on spatial frequency, with the aim of narrowing down on explanatory possibilities.

Subjects

11 subjects participated in the study, all of them in both experiments. They had normal or corrected-to-normal visual acuity (decimal acuity > 1.0), and provided written informed consent. The study was approved by the local review board and followed the tenets of the Declaration of Helsinki.

Stimuli and procedure

Random patterns and natural images were used as stimuli (Figure 1). Pairs of Stimuli were presented sequentially and subjects had to perform a two-interval forced-choice task. Each pair consisted of two instances of the same picture (either the same random pattern or the same natural image), but one was displayed at one third of the linear size of the other, and the maximum spatial frequency (low-pass filter cutoff) in the image differed between the
two pictures. Importantly, image blur, not the subjects’ visual acuity, was the limiting factor for detailedness. Spatial filtering was performed using a wavelet-like transform (see Appendix A).

For equality of objective detailedness, we chose the filter parameters such that the number of spatial-frequency cycles at the respective cutoff frequencies (i.e. the number of details in the image) were identical for both the large and the small stimulus. This was achieved with a cutoff frequency of the small stimulus that was three times that of the large stimulus, i.e. the reciprocal of the linear size ratio. For this case, we predicted that subjects would perceive the small stimulus as being more detailed than the large stimulus. The measurement procedure aimed at compensating this perceptual difference by shifting the cutoff frequencies of the two stimuli in opposite directions (Figure 2). If our prediction were true, it would be necessary to shift the cutoff frequency of the larger stimulus to higher values and to shift the cutoff frequency of the smaller stimulus to lower values. If the opposite of our prediction were true, we should find a reversed result.

Because the Weber law has been shown to appropriately describe many aspects of vision, including spatial frequency discrimination (Campbell, Nachmias, & Jukes, 1970; Mayer & Kim, 1986), we applied frequency shifts that were identical in magnitude on a log scale for both stimuli.

Figure 1. Stimulus pictures used in Experiment 1. Large (top row) and small (bottom row) images were identical except for the size.

Figure 2. Detailedness is controlled via the cutoff frequency of a low-pass filter. Because the linear size of the small stimulus is one third of the size of the large stimulus, the cutoff frequency of the small stimulus needs to be three times that of the large stimulus for equal objective detailedness. If, in case of objective equality, the small stimulus appears more detailed than the large stimulus, a higher cutoff frequency for the large stimulus and a lower cutoff frequency for the small stimulus is chosen to achieve subjective equality, as indicated by the red arrows. This compensatory frequency shift was determined with a staircase procedure. Because the frequency is shifted by a factor of s in both the large and the small stimulus, the total difference in cutoff frequency will be $s^2$ or, on a logarithmic scale as used in the Results section, 2 log s.
large and small stimuli, but in opposite directions. Due to the symmetric nature of the frequency shifts, we use the term “center frequency” for the pair of frequencies that corresponds to equal objective detailedness in large and small stimuli. When reporting the results, we will combine the frequency shifts of the large and the small stimulus into a total frequency shift. All results will be given in terms of the logarithm of the spatial frequency (logSF).

By pressing keys on a keyboard, subjects indicated which of the two stimuli appeared to be more detailed than the other. In case that the subjects had missed a stimulus presentation, they could request a repetition of the trial by pressing a third key instead of one of the two response keys. For the repetition, exactly the same parameters were used as in the original trial. The subjects were given a written explanation of the procedure, which detailed the task and emphasized that the subjects should not attempt to assess individual items in the images, but rather rely on their overall impression.

Igor Pro (Wavemetrics Inc.) was used for stimulus generation and display, response collection, and data analysis. Stimuli were created by a Mac Pro computer and presented on a gamma-corrected black-and-white CRT screen (Philips GD 402) at a distance of 114 cm from the subject at a contrast of 98%. For the large stimuli, the size of the random pattern was 22.1° · 22.1°, the diameter of the images was also 22.1° including the contrast taper (see Figure 1). The dimensions of the small stimuli were one third of these values.

A trial was introduced by an auditory alert, followed by a 1-s period in which the subject could prepare for the stimulus presentation. The two stimuli, large and small, were then presented sequentially in random order, each for 750 ms with an inter-stimulus interval of 80 ms during which the stimulus area was gray. There was no time limit for the subject’s response. Based on the response, the next stimuli were prepared (see Appendix A), which took around 7 s. Each block of the experiments encompassed 31 trials.

The first experiment consisted of eight blocks, including one with random patterns and the remaining seven each using a different photograph. For each trial, a low-pass filter was applied to these “raw” stimuli. The frequency shifts were controlled by a simple up-down staircase procedure (Leek, 2001) in response to the keys pressed by the subject, targeting the frequency difference at which the detailedness of the larger picture subjectively matched the detailedness of the smaller picture. The staircase procedure started at equal objective detailedness with a step size of 0.12 logSF. Over the course of a run, the step size was linearly reduced, reaching 0.017 logSF in the last trial.

The second experiment was identical to the first one in most respects, except that only random patterns were used, and three different center frequencies were compared, corresponding to cutoff frequencies of 1.4, 2.3, and 3.6 cpd in the large stimuli, and 4.3, 6.8, and 10.8 cpd, respectively, in the small stimuli.

### Analysis

For each block of the experiment, the psychometric function was obtained through a maximum likelihood fit assuming a cumulative Gaussian curve shape. Threshold (inflection point) and maximum slope of the curve were determined from the parameters of the fitted curve. 95% confidence intervals were computed through a bootstrap procedure (Good, 2005). For this, 1000 repetitions of the experiment were simulated by drawing trials (with putting back) from the real run of the experiment. For each simulated run, the psychometric function was fitted as described above. The range defined by the 2.5% and 97.5% percentiles of the resulting values for threshold and slope was taken as confidence interval. Non-parametric statistical tests, namely Wilcoxon and Friedman tests, were used for inferential testing since we could not assume the data to follow a normal distribution.

### Results

One of the eleven subjects was excluded from the analysis because he reported immediately after the experiment that he had fundamentally changed his decision criterion during the course of the experiment after realizing that he had misinterpreted the instructions.

#### Experiment 1: Demonstration of the effect with different images

For subjective equality, subjects consistently required the cutoff frequency of the larger image to be higher than that of the smaller image (Figure 3), meaning that the small stimulus was subjectively more detailed in case of equal objective detailedness. This was significant for all images individually (one-tailed Wilcoxon test, see Figure 3 for P values). Due to single outliers, for some images only P values of 0.042 were obtained, i.e. significant on an individual level, but not when a correction for multiple testing was applied (sequential or normal Bonferroni adjustment). The effect size differed significantly between stimulus types in the range of 0.26–0.48 logSF (Friedman test, P = 0.016). To assess the consistency of the results, we performed a principal component analysis, yielding two significant components accounting for a total of 94% of the variance (79% and 15%, respectively). A varimax rotation was applied to these components. The first of the rotated components accounts for the main detailedness effect in eight of the ten subjects. Its high value also implies a consistent pattern of relative effect strengths among subjects. The second rotated component represents the two subjects that produced deviating results, which were anticorrelated between themselves. This is consistent
with what can be seen in Figure 3, where one subject produced an outlying result with the random pattern stimulus, and another subject produced outlying results with most of the other stimuli. As Figure 4 shows, the first component contributes nearly equally to the results of all stimuli, while the second component exhibits a large variability across stimuli.

Experiment 2: Effect size as a function of spatial frequency

In eight out of ten subjects, the effect increased monotonically with spatial frequency (Figure 5). On the group level, the differences between the lowest spatial frequency and both the medium and the high center frequency was significant taking multiple testing into account (Wilcoxon test, both $P = 0.002$). There was no correlation between the compensatory frequency shift and the slope of the psychometric function (Figure 6; Mann–Kendall test, $\tau = 0.11, P = 0.65$).

Because the same subjects participated in both Experiments 1 and 2, we could directly compare two data sets obtained with the medium-frequency random pattern, which was used in both experiments (Figure 7). We found a sizable correlation (Mann–Kendall test, $\tau = 0.62, P = 0.0095$) with a slightly, but not significantly, smaller effect in the second experiment, compared to the first experiment (median, $-0.42 \, \logSF$ versus $-0.48 \, \logSF$; Wilcoxon test, $P = 0.85$). The median absolute difference between experiments was 0.064 logSF.

Discussion

For subjectively equal detailedness, nine out of ten subjects required the small stimulus to have a lower cutoff frequency than the larger stimulus, by 0.4 logSF. This implies that in all these cases subjective detailedness was higher in the small images when the objective detailedness was identical.

Obviously, the present experiment depends crucially on what the subjects consider to represent detailedness. Possibly, the subject for whom deviating results were obtained with most stimuli differed in this respect from the other participants, despite the written instructions that we had provided. But even for the remaining subjects, who performed in a mostly consistent manner, it needs to be considered whether they could have confounded detailedness with other stimulus characteristics. The magnitude of the effect does not provide a hint in this respect. If it were simply absolute spatial frequency, or the size of individual items in the images, that was equalized in the experiment, subjects should have reached a factor of three, i.e. 0.47 logSF. However, for most stimuli the
effect was clearly smaller, and furthermore, the effect of the center cutoff frequency (Experiment 2) would not be explained. On the other hand, it is evident that the subjects did not simply estimate the number of structures in the images, since this number was identical for both large and small stimuli. While it is clear that the detailedness ratings are ultimately rooted in the stimulus properties, additional explanatory steps beyond a simple assessment of spatial frequencies are required.

A first possible explanation, or rather class of explanations, are differences in perceived contrast (cf. Peli, 1990). The modulation transfer function of the visual system, which describes contrast sensitivity as a function of spatial frequency, has its maximum at around 4 cpd (Owsley, Sekuler, & Siemsen, 1983) in the age group of our subjects, who were in their 20s and 30s. Between 2 and 4 cpd, there is only little change in contrast sensitivity. The cutoff frequencies tested here are around or above the maximum of the modulation transfer function. Above the maximum, the contrast sensitivity is lower for the cutoff frequency of the small image than for that of the large image. Therefore, one would expect the small image to appear less detailed than the large image, which is opposite to what our data shows. A second contrast-related factor to take into account is the role of simultaneous blur contrast (Webster, Webster, & Taylor, 2001). A blurred surround makes an image appear sharper. The matting around the smaller image was homogeneously gray and as such had less contrast and, at least in some broad sense, more blur than the peripheral parts of the larger image. This may have led to a less blurry higher-contrast appearance of the smaller image, which may have resulted in a subjectively higher detailedness. We have been unable, though, to find evidence in the literature that a homogeneous field would in fact have the effect of a low-contrast maximally blurred stimulus, as previous studies appear to have used patterned stimuli in all cases for inducing blur adaptation or simultaneous blur contrast.

A second explanation would be the size of the spotlight of attention (Treisman & Gelade, 1980), which may be too small to cover the large images in full. If so, the amount of details that are within the extent of the spotlight...
would naturally be higher for the smaller stimulus. Depending on the interactions between the stimulus size, the size of the spotlight of attention, and the density of details in the stimulus, this might affect the rating of detailedness. A comprehensive assessment of the individual influences of different stimulus parameters may help elucidating whether this explanation can be sustained. Similarly to focal attention, the limited spatial extent of foveal vision and the relatively reduced extrafoveal vision may affect detailedness. However, cortical processing usually leaves us unaware of the reduced acuity in extrafoveal parts of the visual field. The explanation would thus require an additional postulate, namely that the central image characteristics, which are perceived foveally, are not extrapolated to the extra-foveal parts of the stimulus. Otherwise, subjects should subconsciously be able to infer the total detailedness of the image from the local detailedness in the center, i.e., detailedness should not be affected by stimulus size.

As a third explanatory approach, we would like to put forward a visuo-cognitive hypothesis. Irrespective of the overall content of an image, in most cases natural images will have some high-frequency content, for instance sharp edges. The visual system will usually assume the presence of such high spatial frequencies. However, the images that reach the cortex from the retina are limited in their detailedness due to imperfections of the eye’s optics and due to processing taking place in the retina and in the lateral geniculate nucleus. The cortical visual system is “accustomed” to this and is generally prepared to accept that there may be more details present in the real world than represented in the image. However, the coarser the image structure, i.e. the lower the high-frequency cutoff, the less likely it is that this cutoff is due to limitations of the visual system. Thus, if the image contains only low spatial frequencies, the visual system will not assume more details to be present in the real-world scene. On the

Figure 5. The effect of baseline spatial frequency on the compensatory frequency shift. Baseline spatial frequencies are given for the large images on the bottom abscissa and for the small images on the top abscissa. The majority of subjects required a larger frequency shift for higher baseline frequencies. This was significant for the comparison of the low spatial frequency with both middle and high spatial frequencies (‘**’ indicates a family-wise error rate below 0.01).
other hand, if relatively high spatial frequencies are present, the visual system is tricked into believing that there are even higher spatial frequencies in the outside world. Obviously, at present this is largely speculation and will need a more formal Bayesian specification. However, the hypothesis shares some concepts with that put forward by Artal et al. (2004), who suggest that the visual system is able to compensate for its own optical aberrations on a neuronal level.

Another factor to consider is the effect of eye movements. The presentation duration of 750 ms allows for multiple saccades to occur. Larger images may be associated with more and larger saccades. However, this would not necessarily result in subjective detailedness being higher in large images, as there is evidence for both visual memory being preserved across saccades (possibly increasing subjective detailedness due to the accumulation of stimulus information) and visual memory being erased by saccades (possibly reducing subjective detailedness) (Ross & Ma-Wyatt, 2004).

The present results were qualitatively fully consistent across eight of the ten subjects, and mostly consistent in a ninth subject. The variability in magnitude does not come as a surprise, since a number of previous studies have found sizable differences between subjects for various tasks that involve the recognition or judgement of blurred stimuli. For instance, the ability to read blurred optotypes differs greatly between subjects (Coppens & van den Berg, 2004), and the threshold for perceiving a stimulus as blurred shows a large variability (Layton et al., 1978; Wang, Ciuffreda, & Vasudevan, 2006). The comparison between identical conditions of the first and second experiment suggests that the dependence of subjective detailedness on image size is reproducible within subjects. We found no systematic relationship between the compensatory frequency shift and the slope of the psychometric function, and thus no evidence that different effect ranges would be associated with different decision strategies that would have been differentially susceptible to noise as reflected by the slope.

**Conclusions**

We confirmed the cursory finding that small degraded images appear more detailed than their larger, but otherwise identical, counterparts, and found the effect to be sizable and robust. The effect is not adequately explained by relying only on properties of low-level visual processing. Subjective detailedness, as investigated in the present study, may be one building block of a medium-level metric of image quality and as such contributes to filling a

![Figure 6](jov.arvojournals.org) Relationship between slopes of the psychometric functions and the compensatory frequency shift. There appears to be no correlation between both values. The graph shows the data obtained with the medium center frequency in the second experiment.

![Figure 7](jov.arvojournals.org) Test–retest reliability of the detailedness effect. Here, the two identical conditions of the first and the second experiment (random pattern, medium baseline frequency) were compared. There is no systematic difference, and the values are highly correlated. There was one outlier (▽) with an unusually high value of 0.58 in the first experiment.
void that is left open by other metrics of image quality, be it subjective preference, meaning-driven ratings, or measures of basic physical image properties.

**Appendix A**

**Stimulus details**

The photographs were public domain images obtained from the Wikimedia Commons website (http://commons.wikimedia.org/), from which square sections were selected for further processing. Random-noise pictures were created by filling the pixel array (492 pixels × 492 pixels) with binary black-and-white noise. Both large and small stimuli were first created at the size the large stimulus. The small stimulus was then obtained through scaling.

Stimulus pictures were produced by applying a spatial low-pass filter to the raw images based on a wavelet-like transform. This involved convolving the images with a series of Gabor-like two-dimensional wavelets, which were defined by their spatial frequency and their orientation. As opposed to the usual Gabor that use a two-dimensional Gaussian as envelope, a two-dimensional Hanning apodization function (also known as raised cosine) was used in order to limit the spatial extent of the wavelet (cf. Heinrich, Andrés, & Bach, 2007). Convolution of the image with different wavelets resulted in a space-frequency-orientation representation from which the image could be reconstructed by deconvolution. For efficiency reasons, only wavelets with 10 different frequencies and 8 different orientations (i.e., 80 combinations) were employed. The frequencies were equally spaced on a logarithmic scale and ranged from 0.1 cpd to a maximum frequency as prescribed by a staircase procedure, taking into account the subject’s response history.

All computations could be performed within approximately 7 s with Igor Pro (Wavemetrics, Inc.), using a parallelized algorithm that made use of multiple processors. This was fast enough to compute the stimuli ‘on-line’, which allowed for an arbitrary choice of the cutoff frequency as prescribed by a staircase procedure, taking into account the subject’s response history.

In order to avoid edge effects, which may be caused by wrap-around during spatial filtering, the natural images were confined to a circular region with a gradual contrast reduction towards the rim of the picture (see Figure 1) that was achieved by multiplying the image with a two-dimensional Hanning apodization function. This was not necessary for the random patterns. However, for these the value range was clipped such that all the values outside of four times the standard deviation of the image data would be mapped to white and black, respectively. If only the extreme values were mapped to black and white, the gray scale would be very sensitive to statistical fluctuations in the image. Because image statistics depend on the image content, a grayscale clipping based on the standard deviation is unfeasible in natural images. Instead, we used a fixed clipping boundary for all natural images.

**Acknowledgments**

We thank our subjects for their participation. The study was supported by the Deutsche Forschungsgemeinschaft (BA 877/18).

Commercial relationships: none.
Corresponding author: Sven P. Heinrich.
Email: sven.heinrich@uniklinik-freiburg.de.
Address: Sektion Funktionelle Sehforschung, Univ.-Augenklinik Freiburg, Killianstr. 5, 79106 Freiburg, Germany.

**References**


