The fragile edges of block averaged portraits

Taku Taira

Department of Psychology and Neuroscience

April 22, 1999

New York University

T.Taira (1999) *The fragile edges of block averaged portraits*. New York University, Department of Psychology honors thesis. http://psych.nyu.edu/pelli

Abstract

The blocky faces painted by Chuck Close and the block-averaged portraits produced by Harmon and Julesz provide a strong counter to the assumption of scale independence in visual perception. When either Chuck Close's or Harmon and Julesz'z portraits are viewed from a great distance, it is indistinguishable from a unmnanipulated photograph of the subject. As one approaches the portrait it still appears to be a complete three-dimensional face despite the fact that the grid made by the blocks is still visible. Scale independence predicts that when the same portrait is viewed up close, one would continue to see a three dimensional face. However when the portrait is viewed at short viewing distances, the face looses its tree dimensional quality and collapses into a grid of flat marks.

We hypothesize that the transition between percept of the grid and the percept of the face is determined by the relative amount of energy in the grid and in the face. According to this hypothesis, increasing the energy in the grid, independently from the energy of the face, will bias the observer towards seeing blocks. Furthermore changing the contrast of the image will decrease the energy of both the grid and of the face to the same degree, we hypothesize that there would be no change in the observer's bias towards seeing the blocks or the face.

The result of all of the manipulations is a decrease in how readily the face collapses compared to the unmanipulated condition. The collapse of the face is not determined by the relative amount of energy in the block and in the face, instead there is something special about the conditions created by unmanipulated block averaging results in the strongest bias towards seeing the blocks.

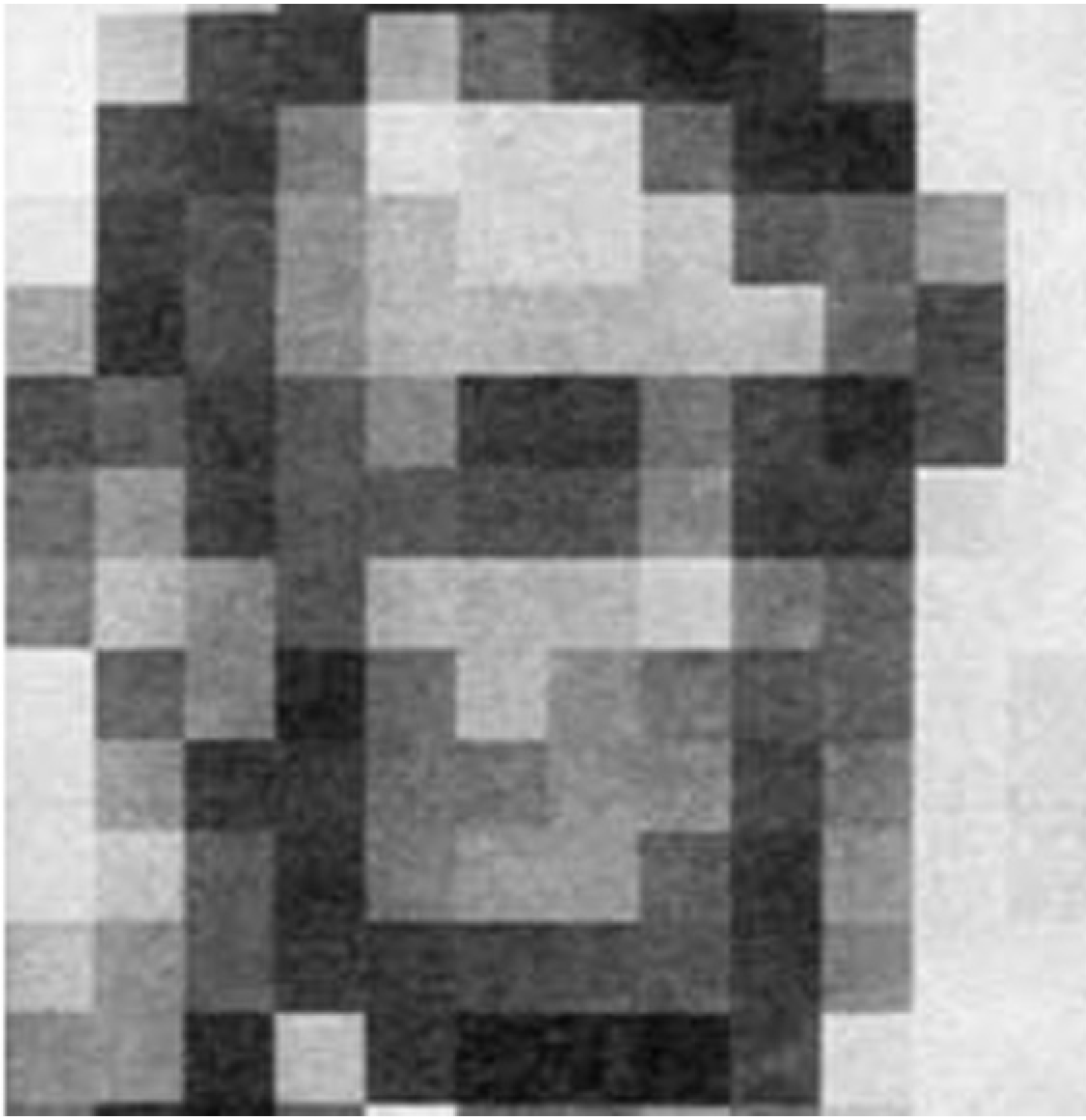
Introduction

It is a long held assumption in the field of visual perception that objects are identified in a manner independent of their size (visual angle subtended by the eye)(Mach 1886). This assumption is only true for objects whose sizes are not so small or far away to be beyond our acuity, or so large or so close to exceed our visual field (Aristotle). This assumption may have arisen because the contrary assumption seems so unlikely. It is hard to imagine that we would identify a friend's face differently when it is two feet in front of us then when it is twenty feet away. When we do such a task, introspection tells us that there is nothing inherently different in recognition of faces at different distances. However, Chuck Close's recent portraits and block averaged portraits reproduced at a large scale provide a strong counter to this assumption of visual scale dependence.

When Chuck Close's later paintings are observed at a distance, they are indistinguishable from a photograph. As one approaches the portrait the face remains three-dimensional even though the individual blocks are visible. However when the same painting is observed at short distances, the three-dimensional face collapses into an array of flat blocks.

The observation of seeing a three dimensional face at one distance and flat blocks at another distance is in direct opposition to the assumption of visual scale independence. The assumption predicts that a three dimensional face would be seen at all distances, because the only thing about the stimulus that changes is the size of it on the retina. This scale dependant phenomenon is powerful and impervious to assumptions and biases. Even if one knows that the painting is of a face, when the painting is viewed up close the collapse of the face cannot be escaped. The masking of the face does not go away because the grid exceeds our acuity and we can no longer see it. We see the three dimensional face over distances where the blocks are still visible but are not able to flatten the face into two dimensions.

This phenomenon of scale dependence shown by the Chuck Close's paintings is not a peculiarity existing only in the art world. In their study of visual perception, Harmon and Julezs (1973), also produced scale dependant portraits, through a process called block averaging. Block averaging is a manipulation where a grid is laid upon a face or any other object. Within each square of the grid, the luminance of the pixels is replaced with the average within that square, resulting in a blocky portrait with uniform squares. Harmon and Julesz's block averaging technique is akin to Close's method of creating his blocky portraits. Chuck Close produces these block portraits by first laying identical grids upon both the canvas and on the photograph that he is painting from. Instead of simply copying the fine details of each square from the photograph to the canvas he gives each square a structure that is inconsistent with the detail in the photograph.

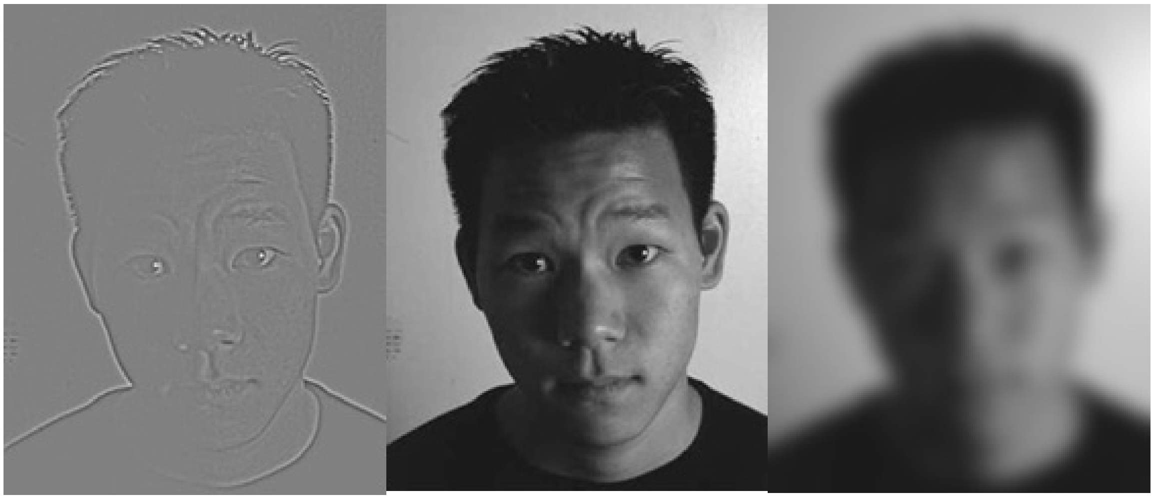


(fig 1) This is a block averaged portrait of a famous person. Can you figure out who it is? Try moving away.

Unlike Close, Harmon and Julesz attempted to not only understand the visual mechanisms that mediated the masking but to also describe them. In their paper "Masking in Visual Perception" Harmon and Julesz block averaged a portrait of a famous historic figure and asked people if they could identify who it is. When this block portrait is first observed, it is very difficult to make out who it is. They reported in their paper what were the manipulations that would improve the recognizability of the block averaged portrait. They found that the identification of the portrait is greatly enhanced by blurring through the methods of squinting, defocusing, moving away from the object and rapidly moving the image or one's head. (figure 4)

Harmon and Julesz'z observations of their portrait are consistent with people's description of viewing Chuck Close's paintings. According to Harmon (1973) "Viewed from close up, these 'block portraits' appear to be merely an assemblage of squares. Viewed remotely, from a distance of 30 to 40 picture diameters, faces are perceived and recognized." Harmon and Julesz propose that the ability for the blocks to obscure the face is mediated by critical band masking.

Critical band masking is the phenomenon that occurs in both audition (Zwicker 1957) and vision (Stromeyer and Julesz, 1972) where a signal presented in noise is most effective at obscuring a signal when the frequency of the noise is similar to the frequency of the signal. In audition, frequency corresponds to the pitch of a sound. The sound of a fire alarm is an example of a high frequency sound while a foghorn would be an example of a low frequency sound. In vision, frequency corresponds to the fineness of detail. The high frequency information of an object would be its edges while the low frequency is the general shape of the object. (figure 2)Radio static is auditory noise and visual static is television noise.



(figure2)
The left image is the high frequency information of the center photograph obtained by high-pass filtering. The right most image is the low frequency information of the center photograph obtained by low-pass filtering.

The critical band theory is based upon the idea of channels. The idea is that an early level of perception, our visual or auditory field is filtered through a parallel group of channels. Each channel is sensitive to only a band of frequencies. A signal is detected by a single channel that is centered on the signal's “critical frequency." The critical band can be revealed by masking the signal with noise of varying frequency. The frequency of noise, which greatly reduces the identifiability of the signal to the greatest degree, corresponds to the signal’s critical band. (figure 3)

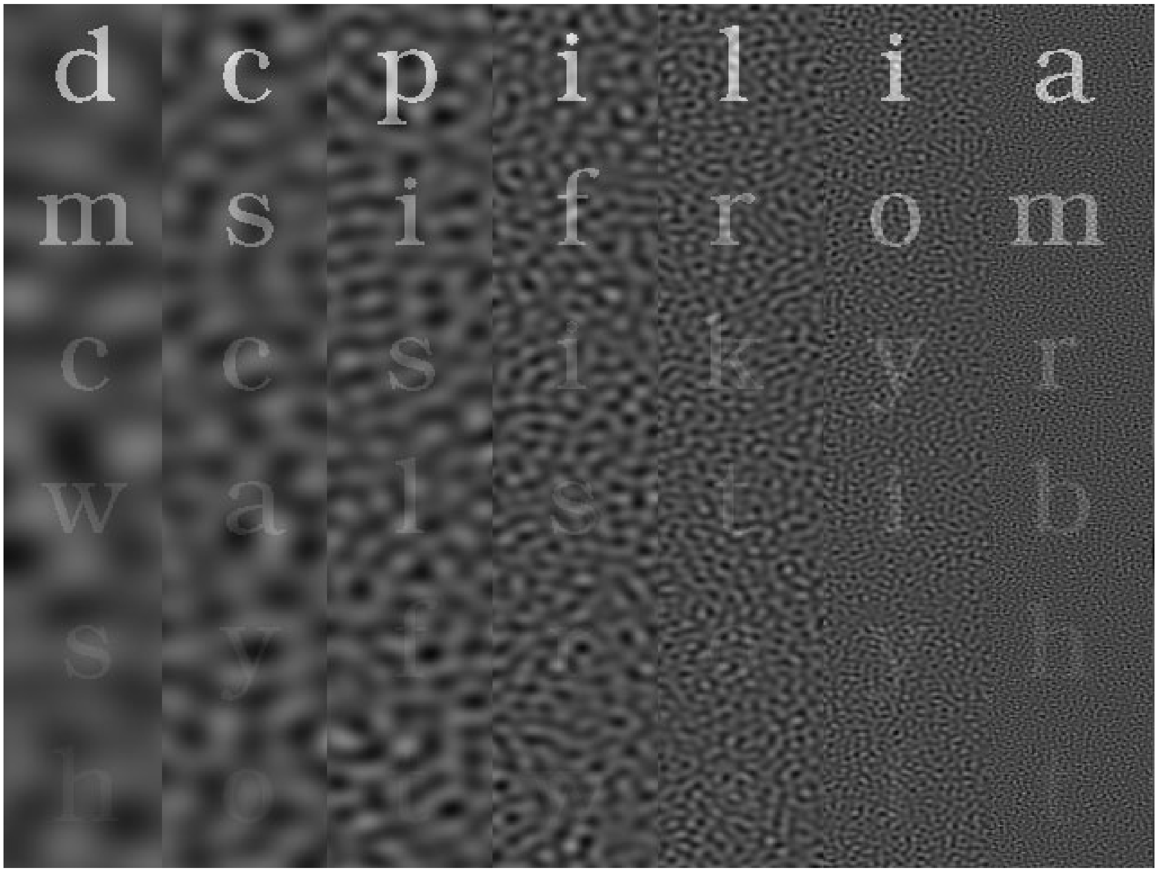


Figure 3
From left to right each column contains noise of higher and higher frequency. From top to bottom are letters that are decreasing in their contrast. Effectiveness of the mask is determined by how far down one can identify in a column. The reason you cannot read very far down the center column is because the critical band for identifying letters is sensitive to the frequency of the noise.

Harmon and Julesz suggested that there is a critical band for face identification. They attributed their masking effect of block averaging to the introduction of “noise” within the critical band for face recognition. To support this conclusion they removed a band of noise whose frequencies are adjacent to what they referred to as the critical band for the recognition of the face. (Figure 4) They found that this manipulation improved the recognition of the face. (Figure 4) Their conclusion was further supported by their observation that removing a similarly sized band of high frequency, far from the critical band of the face has no effect on the identifiability of the face.

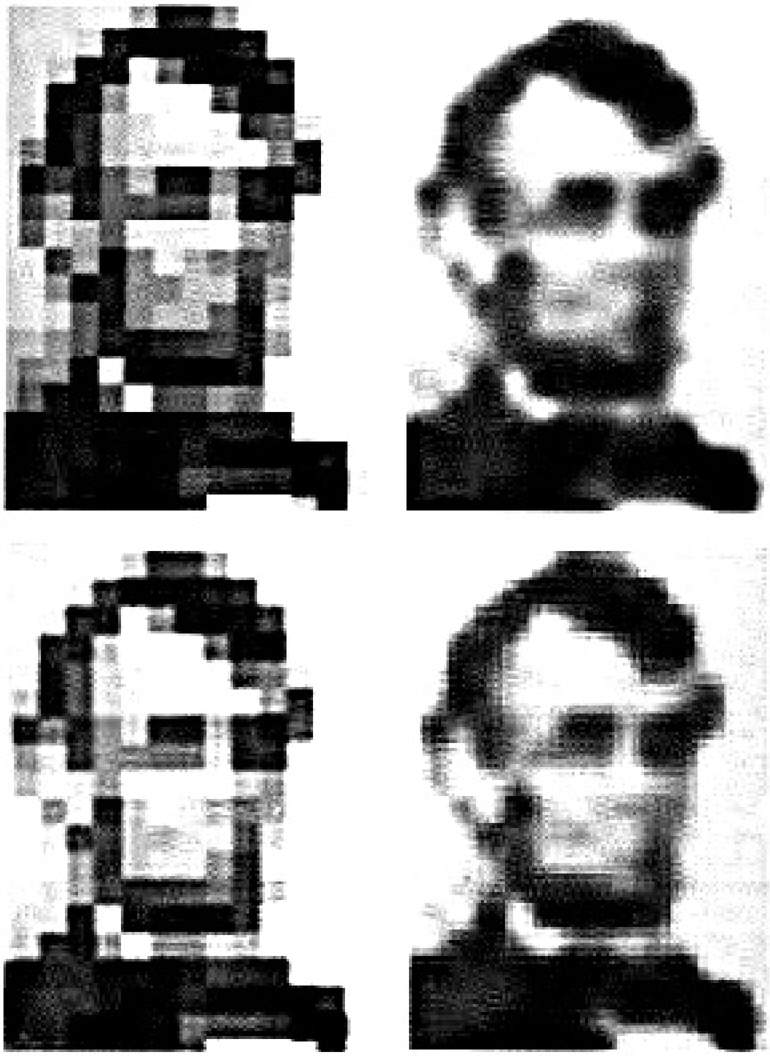


figure 4
The top left image is the untouched block portrait
Top right is the portrait that is more identifiable through blurring
Bottom right has the high frequencies removed with no enhancement of identifiability
Bottom left has the frequencies near the critical band removed with an enhancement in identifiability

Even though the critical band explanation works at first approximation, it falls apart under close scrutiny. Pelli (1999) also studied the recognition of blocky objects. Instead of studying the blocky objects of Harmon and Julesz, Pelli studied the recognition of the of Chuck Close's portraits. Instead of asking the question, "can you recognize this as a face," Pelli used a psychophysical "nose test" to quantify the strength of the masking. Pelli quantified the observer's bias towards seeing the face and seeing the blocks by asking Observers to concentrate on the bridge of the nose and find the viewing distance at which they saw a transition between a three-dimensional to a two-dimensional nose. The bridge of the nose is emphasized because Close paints the nostrils with a detail greater than the rest of the portrait. Furthermore, Pelli describes this transition in his paper. “As the viewer approaches [the portrait], the nose suddenly collapses, as though surgically excised, leaving a flat patch of skin where the nose had been... the transition is abrupt.”

Harmon and Julezs’ critical band explanation predicts that the “nose test” would reveal a fixed number of marks across the face independent of face size. Keeping the marks across the face constant as face size changes will maintain the relationship between the frequency of the blocks and the frequency of the face. However Pelli found that as the size of the face increases, more blocks are needed to see the face. With increased number of blocks across the face, the spatial frequency of the grid will also increase apart from the spatial frequency of the face. Therefore Harmon and Julesz’s critical band masking cannot explain this phenomenon.

At first glance, this comparison of Pelli and Harmon and Julezs may seem like comparing two completely different things. However, assessing dimensionality is a very good correlate to the loss of the ability to recognize the face. People looking at both the Harmon and Julezs Lincoln and Chuck Close portraits describe the sensation of losing the perception of the face occurs at the same distance when they see a transition from the face to flat marks. Furthermore, Chuck Close himself mentions that when he is painting, an arms length from the canvas, he cannot identify the face which he is painting. Identifiability and dimensionality seem to be part of the same mechanism.

Since the competition between the block mask and the face cannot be explained by critical band masking, another explanation could be that the phenomenon is mediated by the relative amount of energy in the blocks and in the face. If the competition is mediated by the relative amount of energy in the edges and face, we should then be able to change the strength of the masking by changing the ratio of the energies. We predict that increasing the energy in the blocks should bias the observer towards seeing the blocks and decreasing the energy in the block should bias the observer towards seeing the face. Similarly, if we were to change the strength of the blocks and the face by equal amounts by changing the contrast of the entire image, we do not expect that there would be any change in the masking.

The transition between the percept of the grid and the face is a change in dimensionality. The competition between the grid and the face are therefore competing for control of the mechanism of extracting depth. The most salient cue for depth in blocky objects is the cue of shape from shading. Shape from shading is the process by which the visual system extracts three-dimensional information from the shading of an object.

Even though shading is a useful cue for depth, Ramachandran (1989) noted that shape from shading is not very strong when it is the only cue for depth. As a result is very shape from shading is dependent upon the edges that bound the shading. He demonstrates this by having two objects with identical shading but with different percepts caused by the difference in the shading boundary. (Figure5)

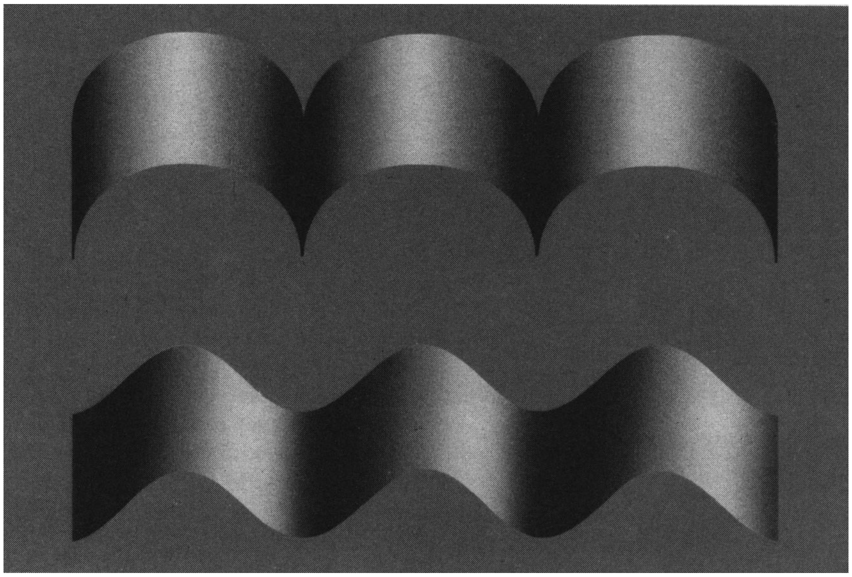


Figure 5
These two images are identical in their shading information but only differ in their boundary edge.
Despite being so similar the percept of these images is very different.
The upper image looks like three cylinders next to each other while the bottom shape looks like corrugated metal.

Ramachandran concludes that, "when shading cues are ambiguous, information from borders helps to resolve ambiguity throughout the image." Perhaps at long viewing distances, when a face or nose is perceived, the borders within the faces and nose are disambiguating the shading information. However, at close distances, the edges of the grid is disambiguating the shading information and causes the percept of the grid.

Methods

Subjects

All subjects in this study have normal or corrected to normal vision. All subjects were kept naive to the predictions of the experiment.

Stimuli

To produce the portraits for the stimuli photographs were taken of several people against a bare white background. All photographs were full frontal. No effort was made to obscure differences in hair. All the photographs were taken in a studio on a Canon A2E, on Fuji Provia slide film, and light with dynalite studio strobes.

To maximize the shading information in the portrait, the models were lit from the side. The strobes were used without reflectors, and were positioned to at an approximately 45-degree angle from the model. The result is a portrait with distinct side lighting with a great deal of shading information. (Figure 6)

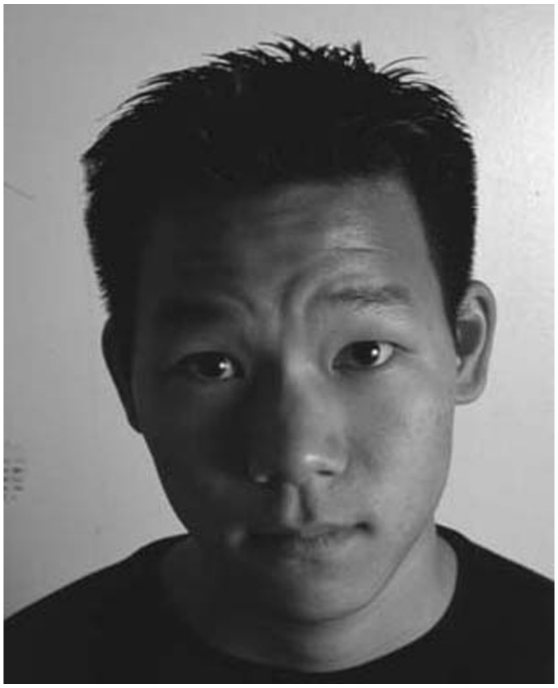


figure 6

All image manipulations were done on an Apple Power Macintosh G3-233 personal computer. The resulting slides were scanned into the computer (Nikon Super Coolscan) at a resolution of 75 dpi. The resulting image is 322pixels by 418pixels. The images were then block averaged in Adobe PhotoShop using the Mosaic filter. This filter performs the Harmon and Julezs block averaging manipulation. The images were then block averaged with square sizes of 10, 20, 30, and 40 pixels. The resulting portraits were printed on a black and white laser printer resulting in portraits with block sizes of .25", .5", .75", and 1".

To extract the edge information separate from the face information we first separated the portrait into its frequency components filtering the portraits through a pyramidal filter. (Figure 7) Pyramidal filtering is a type of filtering which separates the spatial frequency components of an image by filtering the image through filters of varying size.

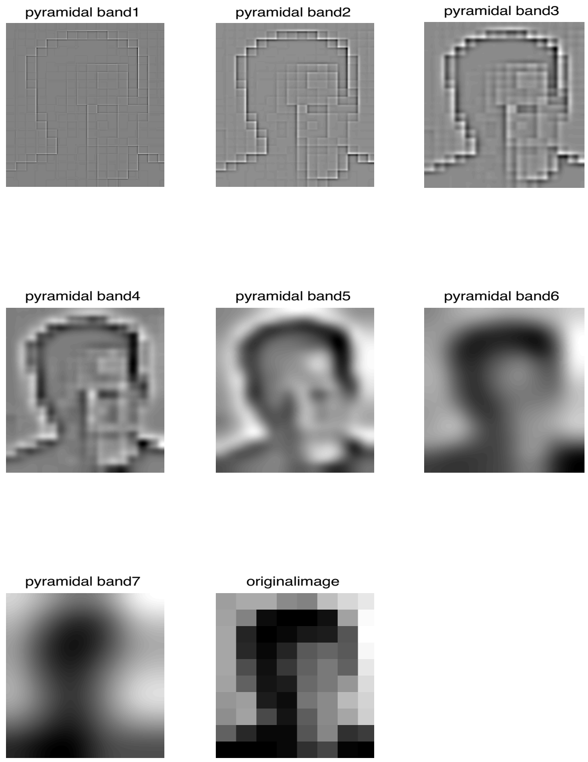


figure 7

Pyramidal filtering was used instead of other filters to avoid any artifacts such as filter ringing which would appear as rings of light and dark separate unrelated to the filtered image . The result of the filtering is, the portrait is separated into seven decreasing levels of spatial frequency. By looking at the figure one can clearly see that at the early levels of the pyramid there is mostly the edges of the blocks and at the lower levels of the pyramid there is no sign of the edges of the blocks.

For each of the block portraits we reconstructed the low frequency levels of the pyramid which did not contain any of the edges of the blocks. The resulting image resembles a blurry version of the original portrait.

(Figure8)

The blurry image was then subtracted from the block portrait. The result is the edge information by itself. The energy in the edges was then manipulated by multiplying or dividing the image by 2. The result is the contrast in the edges is doubled or halved this corresponds to the energy being increased by 4 or reduced by 1/4. (energy is contrast squared) This manipulated edge was then added back to the blurry image. The result is one stimulus with the ratio of energy quadrupled, quartered, and unmanipulated.

To reduce the contrast of the image we filtered the images through a pyramidal filter. We then extracted the luminance information contained in the lower levels of the pyramid (level 7). We then either added or subtracted this luminance information from the blocky portrait. Subtracting the luminance information results in a lower contrast image while adding the luminance information increase the contrast. The result is three stimuli, high, low, and un-manipulated contrast. (Figure 8-12)

Procedure

The stimulus is presented at approximately eye level. The subject is then instructed to move back and forth and find the point at which they see a sharp transition between a three-dimensional and 2-dimensional. Observers were asked to concentrate on the bridge of the nose, to make this task comparable to Pelli's "nose test". The distance from the eye of the observer to the picture is measured. This point is called the observer's critical distance. The critical distance was measured from the eye of the observer to the stimulus, for each of the 5 conditions (un-manipulated, doubled energy, halved energy, high contrast and low contrast.), both faces, and two block sizes for each face. Within each face and block size, each condition was presented in random order. The observers were kept naive to the condition they were observing at each instance, to prevent bias from expectations. Testing lasts between 15 and 20 minutes.

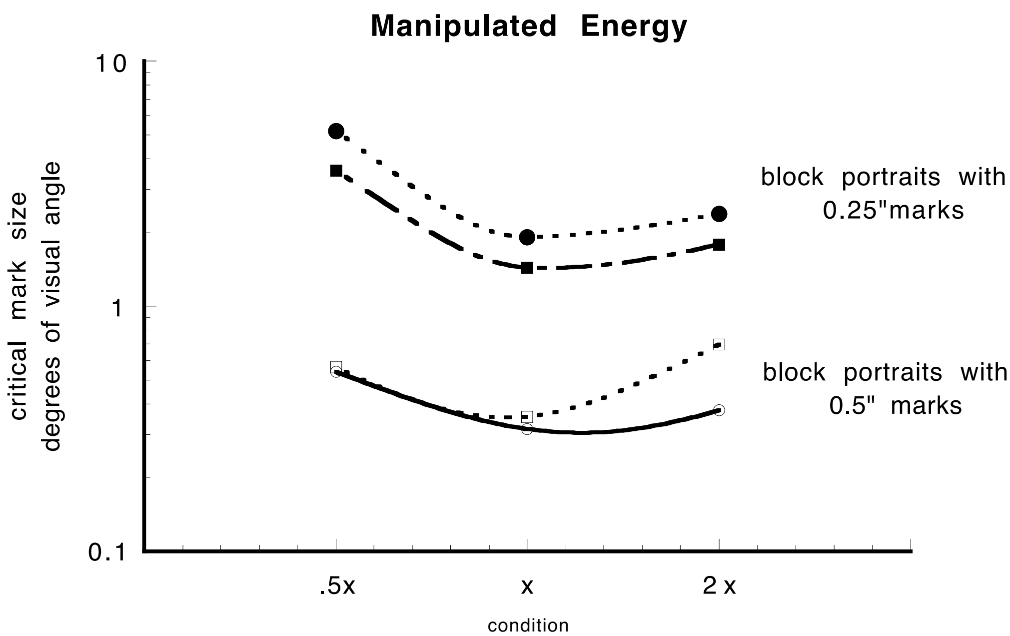
Each critical distance was converted to degrees of visual angle subtended by a single mark. This measurement is called the critical mark size. This conversion allowed us to compare data between different block sizes. The measurement of degrees of visual angle is also more accurate measurement of the stimulus at the level of the retina, than distance. Critical distance is converted to degrees of visual angle by calculating the arctangent of the size of the marks divided by the critical distance.

Data from the stimuli with .75" and 1" blocks could not be collected because observers were unable to ever see the nose, even at very long distances. Data from the face with a block size of .25” is questionable because observers reported difficulty in getting close enough to the stimulus to make it collapse.

Results

Observers describe the stimulus with doubled edge contrast as being the most block like and the stimulus with halved edge contrast as being the most face like. At first approximation, this introspection seems to confirm our hypothesis

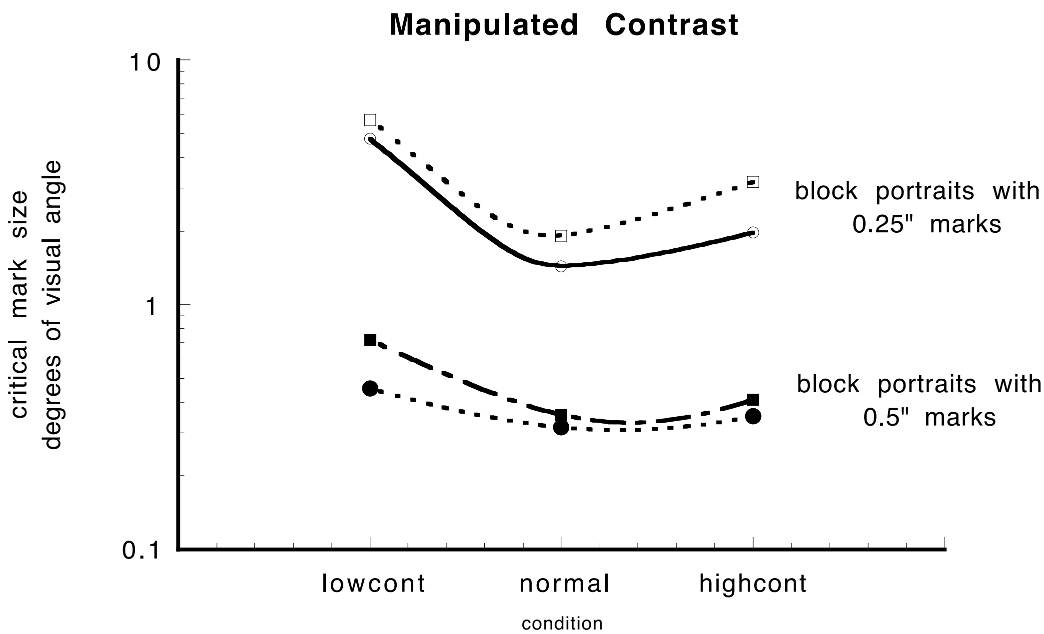
The data for two observers in the manipulated energy task is shown in figure 13.



The vertical scale is critical mark size in degrees of visual angle.
The top curve is data for two different observers for block portraits with mark size of 0.25".
The bottom curve is the data for the same two observers for block portraits with mark size of 0.5"
.5x is the condition where the contrast of the edges has been halved
x is the unmanipulated condition
2x is the condition where the contrast of the edges has been doubled

The data is plotted in units of critical mark size, for each stimulus condition. We expected that decreasing the energy of the edges would lead to the largest mark size and increasing the energy of the edges would lead to the smallest mark size. Small mark size corresponds to larger viewing distance which in turn corresponds to a observer's bias towards seeing blocks. However in all four curves the unmanipulated condition has the smallest critical mark size. In all four curves both manipulations increase the critical mark size corresponding to a decrease in the masking by the grid. This trend is consistent across stimuli and observers.

The data from the manipulations of the overall contrast of the image show the same trend. Figure 14 shows the data for the same two observers comparing the normal condition with the increase and decrease of the overall contrast of the image.



The vertical scale is critical mark size in degrees of visual angle.
The top curve is data for two different observers for block portraits with mark size of 0.25".
The bottom curve is the data for the same two observers for block portraits with mark size of 0.5"
Lowcont- is the condition where the overall contrast has been decreased
Normal- is the condition where the contrast has not been changed
Highcont- is the condition where the contrast has been increased.

As with the data from the manipulated edge energy, the condition with the greatest amount of masking shown by the smallest mark size is the unmanipulated condition. As with the manipulated edge energy, increasing and decreasing the overall contrast of the image decreases the masking, shown by the increase in the mark size. This trend is consistent between observers and stimuli.

Conclusion

The data obtained in this experiment is surprising and mysterious. The data clearly show that our hypothesis is wrong. The phenomenon of scale dependence in viewing blocky portraits cannot be due to the relative amount of energy in the face and the grid. Instead we conclude that there is something special about the conditions that arise from unmanipulated block averaging. This unmanipulated condition causes the observer to have the strongest bias towards seeing blocks. However these conditions are fragile. For some reason all of the manipulations cause a perturbation in the optimal conditions created by the masking.

The most surprising result is the effect of changing the contrast of the image. This manipulation should be identical to producing a high contrast and low contrast version and unmanipulated version of the face before block averaging. We know from our previous studies that critical mark size does not vary under these conditions. Our manipulation of contrast is after the block averaging and somehow changes the optimal conditions created by block averaging.

This phenomenon cannot be explained by critical band masking, nor can it be explained by the relative amount of energy in grid and face. We do not have a satisfactory explanation for this result. We are truly baffled.

Acknowledgements

I would like to thank my sponsor Denis Pelli for his guidance and input. I would also like to thank Melanie Palomares, Sarah Bassin, Ted Coons, and Breena Miller for their helpful suggestions on writing and for being supportive. I would also like to thank all the members of my honor's seminar class and our guide Jim Matthews for their feedback on my experiments.

References

Aristotle, Poetics translated by J. Hutton. New York: Norton, chapter 7 1982

Harmon L. (1973) Identifying Faces, *Scientific American*, 229 (5), p.71

Harmon L and Julesz B. (1973) Masking in Visual recognition: Effects of Two-Dimensional Noise. *Science*, 180, 1194

Mach, E. The Analysis of Sensation, translated by C.M. Williams. London: Routledge Press, p.109

Pelli, D. G. (1999) Chuck Close's demonstration of the size dependence of vision, in submission

Ramachandran V. S. (1988) Perceiving Shape From Shading, *Scientific American*, August, 76-83

Stromeyer, C.F. and Julesz, B. (1972) Spatial frequency masking in vision: critical bands And spread masking. *Journal of the Optical Society of America*, 62, 1221-1232

Zwicker, E. (1957) Low Level Noise Masking, *American Journal of Acoustic Research*, 29, p.548