Pergamon

PII: S0042-6989(97)00052-7

Vision Res., Vol. 37, No. 18, pp. 2609–2629, 1997

Printed in Great Britain
0042-6989/97 $17.00 + 0.00

Capture and Transparency in Coarse Quantized Images

M. CONCETTA MORRONE,* DAVID C. BURR*†‡

Received 16 February 1996; in revised form 24 September 1996

This study examines the effect of coarse quantization (blocking) on image recognition, and explores possible mechanisms. Thresholds for noise corruption showed that coarse quantization reduces drastically the recognizability of both faces and letters, well beyond the levels expected by equivalent blurring. Phase-shifting the spurious high frequencies introduced by the blocking (with an operation designed to leave both overall and local contrast unaffected, and feature localization) greatly improved recognizability of both faces and letters. For large phase shifts, the low spatial frequencies appear in transparency behind a grid structure of checks or lines. We also studied a more simple example of blocking, the checkerboard, that can be considered as a coarse quantized diagonal sinusoidal plaid. When one component of the plaid was contrast-inverted, it was seen in transparency against the checkerboard, while the other remained "captured" within the block structure. If the higher harmonics are then phase-shifted by π, the contrast-reversed fundamental becomes captured and the other seen in transparency. Intermediate phase shifts of the higher harmonics cause intermediate effects, which we measured by adjusting the relative contrast of the fundamentals until neither orientation dominated. The contrast match varied considerably with the phase of the higher harmonics, over a range of about 1.5 log units. Simulations with the *local energy* model predicted qualitatively the results of the recognizability of both faces and letters, and quantitatively the apparent orientation of the modified checkerboard pattern. More generally, the model predicts the conditions under which an image will be "captured" by coarse quantization, or seen in transparency. © 1997 Elsevier Science Ltd

Spatial illusion Transparency Image recognition Feature detection

INTRODUCTION

When images are *coarse quantized* by setting all pixels within regular blocks to the average level, they become completely unrecognizable, although there remains sufficient information at low spatial frequencies for recognition (readily verified by blurring the image). This now classic illusion, first demonstrated by Harmon (1973; Harmon & Julesz, 1973), is one of the most compelling of visual effects. Examples of the blocking phenomenon are now common in most vision textbooks, and are often used as a device for artists, for special effects in publicity, and to disguise identity on television (not very effectively as the identify can be unmasked by simple blurring). However, despite the many demonstrations that coarse quantization reduces recognizability drastically, there have been very few serious attempts to quantify the strength of the phenomenon.

Costen *et al.* (1994, 1996) have measured reaction times and accuracy for recognizing coarse quantized faces. Their results show that reaction times increase and accuracy decreases with decreasing sampling rate, at a higher rate than blurred images with the same information content. However, the results measure only per cent correct performance at a given difficulty level, without attempting to establish a more quantitative measure of performance, such as a contrast or signal-to-noise threshold. Another difficulty is that the images used in their studies were displayed abruptly, a situation that favours low spatial frequencies (e.g., Burr, 1981). Uttal *et al.* (1996a, b) have also made measurements along these lines, using briefly presented stimuli, and small images. The effects of blocking were quite pronounced, although thresholds were not measured.

The original explanation of Harmon & Julesz (1973) for the phenomenon was that mechanisms tuned to the high spatial frequencies introduced by the blocking *mask* the low spatial frequencies that contribute to recognition. However, this explanation has been questioned, both on the grounds that the power of the spurious frequencies is low compared with the image spatial frequencies, and because of experimental evidence showing that recognizability can be restored by *adding* further high-frequency

*Isituto di Neurofisiologia del CNR, Via S. Zeno 51, Pisa 56127, Italy.
†To whom all correspondence should be addressed at Pisa address [*Tel*: +39 50 559719; *Fax*: +39 50 559725; *Email*: dave@in.pi.cnr.it].
‡Dipartimento di Psicologia, Università di Roma "la Sapienza", Via dei Marsi 78, Rome, Italy.

noise, at orientations different from the spurious frequencies (Morrone *et al.*, 1983). Several other explanations suggest a more active role of the spurious blocking components.

The effects of blocking on recognition are well explained by David Marr's (1982; Marr & Hildreth, 1980) theories of the construction of the "primal sketch" of an image. Marr assumes that a feature will be marked in the primal sketch only if there exists a correspondence in position and orientation of zero crossings over a continuous range of operator sizes. For blocked images, the zero-crossings that define the face occur at similar positions to those of the medium scales (which follow a grid-like structure), so the features marked for the primal sketch are mainly those of the higher channel, that do not form the impression of a face. An important assumption to this theory is that we are subjectively aware of the primal sketch, but not of the zero-crossings from which it is made (Marr, 1982, p. 73), so we do not have independent access to the low-frequency information of the blocked images (although low-level hardware does sense this information). Canny (1983, 1986); Witkin (1983) and Koenderink (1984) have developed this idea further, proposing different ideas for continuous integration of features across scales. Canny (1983) extended his model, which synthesises edge information by analysing different scales progressively from high to low, to simulate the blocking illusion. The simulations predicted both the illusion, and the improvement of recognizability of the blocked image after corruption by random noise.

The MIRAGE algorithm of Watt & Morgan (1985) also predicts the blocking illusion, as well as the breakdown of vernier and stereo acuity with sampling (Watt & Morgan, 1982; Morgan & Watt, 1984). An essential part of their model is the obligatory recombination of spatial frequency channels (after an early rectification) before object recognition, accounting for the Lincoln illusion of Harmon and Julesz, and other effects of sampling.

We have also suggested a similar explanation, based on the *local energy* model of image organization (Morrone & Burr, 1988, 1993; Burr & Morrone, 1990, 1992, 1994). This explanation also assumes that vision must construct a symbolic "feature-based" description of the image, not unlike the primal sketch. As with Marr's model, the features are calculated separately at each scale (by peaks in local energy rather than zero-crossings), but then recombined (rather than requiring correspondence across adjacent scales). For recombination, the separate feature maps are each given an indetermination of localization factor, proportional to scale size, then summed. This automatically favours the high-scale operators that respond to the high-frequency block structure, so this dominates the combined map. The low-scale information therefore becomes unavailable as an independent source of information (Burr & Morrone, 1990, 1994; Morrone & Burr, 1993).

An important concept for the local energy model is that image features occur at points of *phase congruence* of the component harmonics. This leads to the prediction that breaking phase congruence across spatial scales should disrupt the appearance of a single feature set, and allowing for multiple features to be perceived nearby, *in transparency*. In this paper we verify this prediction by varying the phase of the spurious harmonics of blocked images. This process restores recognizability in blocked images, causing the low-frequency image to be seen in transparency. We study this effect further with a highly simplified blocked image derived from a checkerboard pattern, and go on to search for a general description of the conditions that support image transparency. These results have been previously published in abstract form (Morrone & Burr, 1994, 1995).

RECOGNITION OF COARSE QUANTIZED FACES

Methods

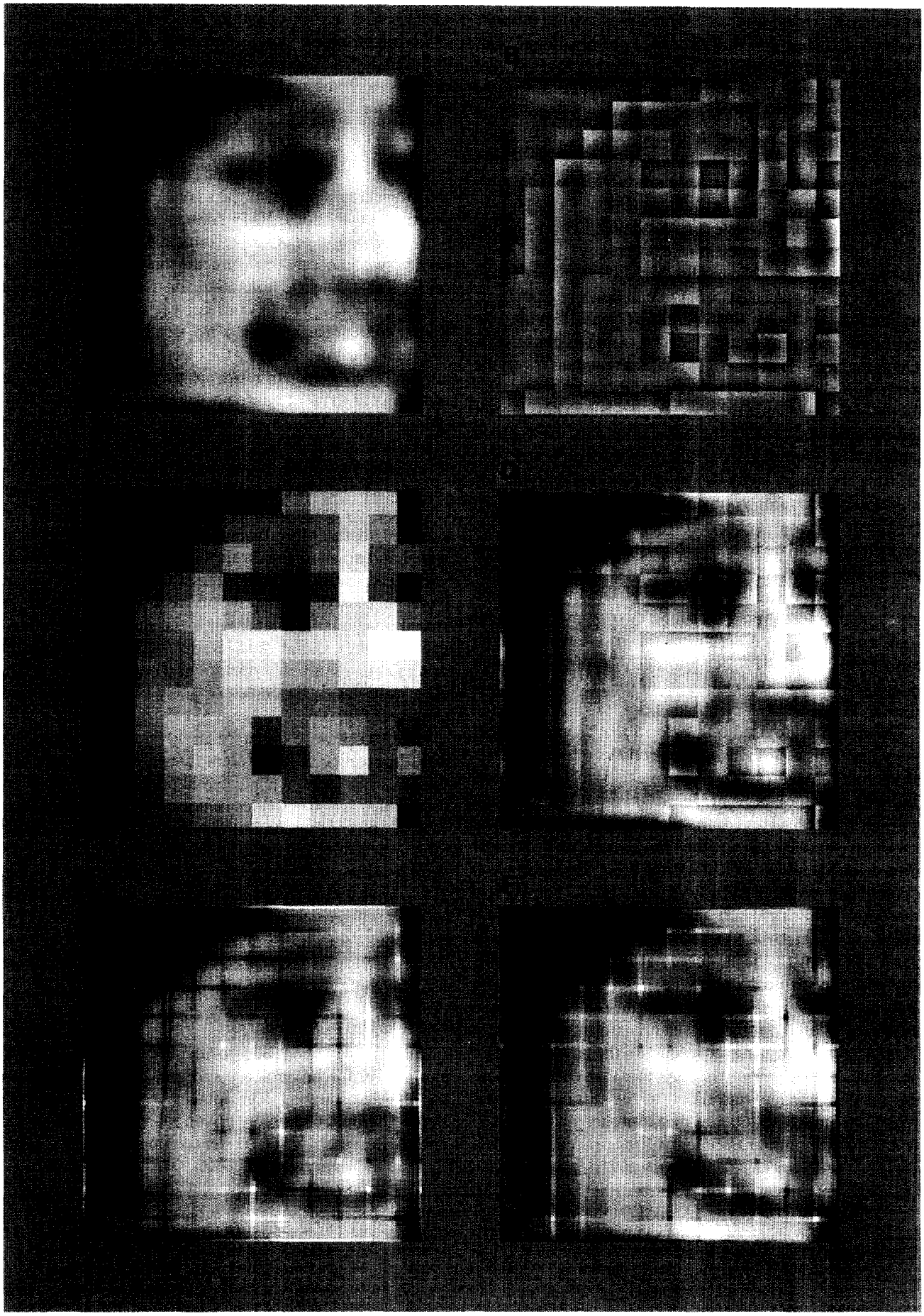
As the original demonstration of Harmon (1973) used a human face (Abraham Lincoln), we first measured the effect of blocking on face recognition, and then the effect of phase.

Stimuli

The faces of six young women, all well known to the subjects were photographed in ten different poses, ranging from full frontal to half-profile in each direction, and with various expressions. All photographs were taken with the same beret to disguise obvious differences such as hair length and colour. To minimize further the use of local rather than recognition cues, all images were mirror-inverted symmetrically around the midline, to provide another 10 poses. Observers were required to identify the face on a particular trial by pressing the appropriate response button (from a choice of six).

All image manipulations were performed on a Silicon Graphics Iris-35 computer, using the HIPS image processing package (Landy *et al.*, 1984). The photographs were scanned (Epson GT-6000 scanner) and digitized to 128×128 pixel images, then scaled to use the full range of 256 grey levels. They were then coarse

FIGURE 1 *(facing page)*. Examples of various versions of the blocked images used for the face recognition study. (C) shows the normal blocked image, obtained by first blurring the image at the Nyquist frequency, then coarse quantizing into 12×12 blocks, and rescaling to use the full luminance range that would not saturate during the phase shifts. This produced an averaged RMS contrast of 0.15, after combination with noise. The other images were all derived from this. (A) is a lowpass filtered version, showing all frequencies below 8.2 c/picture width (ideal filter). (B) shows the spurious high frequencies, obtained by subtracting (A) from (C). (D) shows the blocked image with the spurious components phase-shifted by π [equivalent to subtracting (B) from (A)]. (E) and (F) show the blocked image with the spurious components phase-shifted by $\pm \pi/2$, following the procedure outlined in Fig. 2. For each of the phase-shifted conditions [(D), (E) and (F)], the lowpass face tends to emerge in transparency behind grid.



quantized to 12×12 blocks [Fig. 1(C)], and lowpass filtered again (ideal filter, 8.2 c/picture width) to produce the blurred images of Fig. 1(A). The spurious frequencies [Fig. 1(B)] introduced by the blocking process were obtained by subtracting the blurred image from the blocked image. The phase of the spurious frequencies was shifted by various amounts (see below) then resummed to the blurred image. Figure 1(D) shows examples of a phase shift of π. Figure 1(E) and (F) show examples with $\pi/2$ and $-\pi/2$ phase shifts.

Two-dimensional phase displacement. To study the effects of high-frequency phase we required an image transformation that changed the high-frequency edges to a different feature, without changing its position, orientation or local RMS contrast, and leaving the amplitude spectrum unaffected. To displace phase by π (changing the sign of all edges) it is sufficient to invert the contrast of the spurious harmonics. For other intermediate phases, however, this is more difficult. Although for one-dimensional (1-D) images this can be simply achieved by applying the Hilbert to the original image (shifting the phase spectrum by $\pi/2$ and multiplying the amplitude spectrum by the sign function), the Hilbert transform is not defined in two dimensions. One possibility would be to apply the Hilbert transform along one prevailing orientation (say vertical), but this would transform only vertical edges to lines, leaving horizontal edges unaffected. Applying along two or more orientations separately, then summing the result would annul part of the amplitude spectrum (because of the inherent symmetry of opposing quadrants of the two-dimensional Fourier Transform).

We therefore devised a transform that insured a local phase shift in two dimensions, to convert odd-symmetric edges to even-symmetric lines of the same local RMS contrast. We took advantage of the regularity of the spurious harmonics and of the fact that the distribution of power is mainly along the horizontal and vertical orientations [see Fig. 2(B, D)], and treat the patterns on a column-by-column and row-by-row basis. The Hilbert transform was calculated for each row of the original image [see sketch in Fig. 2(E)], by adding $\pi/2$ to the phase spectrum of the input signal and multiplying the amplitude spectrum by the sign function, and then by applying the inverse Fourier transform. A new image was built [Fig. 2(F)] from the Hilbert transform of the single rows. Applying the same procedure to the new image, we then calculated the Hilbert transform for each column. The resulting image was added to the original lowpass frequency image to obtain the stimulus phase-shifted by $\pi/2$. It is important to note that this procedure affects only the phase spectrum, leaving the amplitude spectrum intact [compare, for example, Fig. 2(B) and Fig. 2(D)].

To obtain phase shifts other than 90 deg, the original and transformed images of the spurious spatial frequencies were combined in appropriate proportions. The output image O(x,y) for a phase shift of ϕ was given by:

$$O(x,y) = \cos(\phi)I(x,y) + \sin(\phi)T(x,y)$$

where $I(x,y)$ is the original image and $T(x,y)$ the transformed image. This procedure affects neither the position nor the local RMS contrast of features.

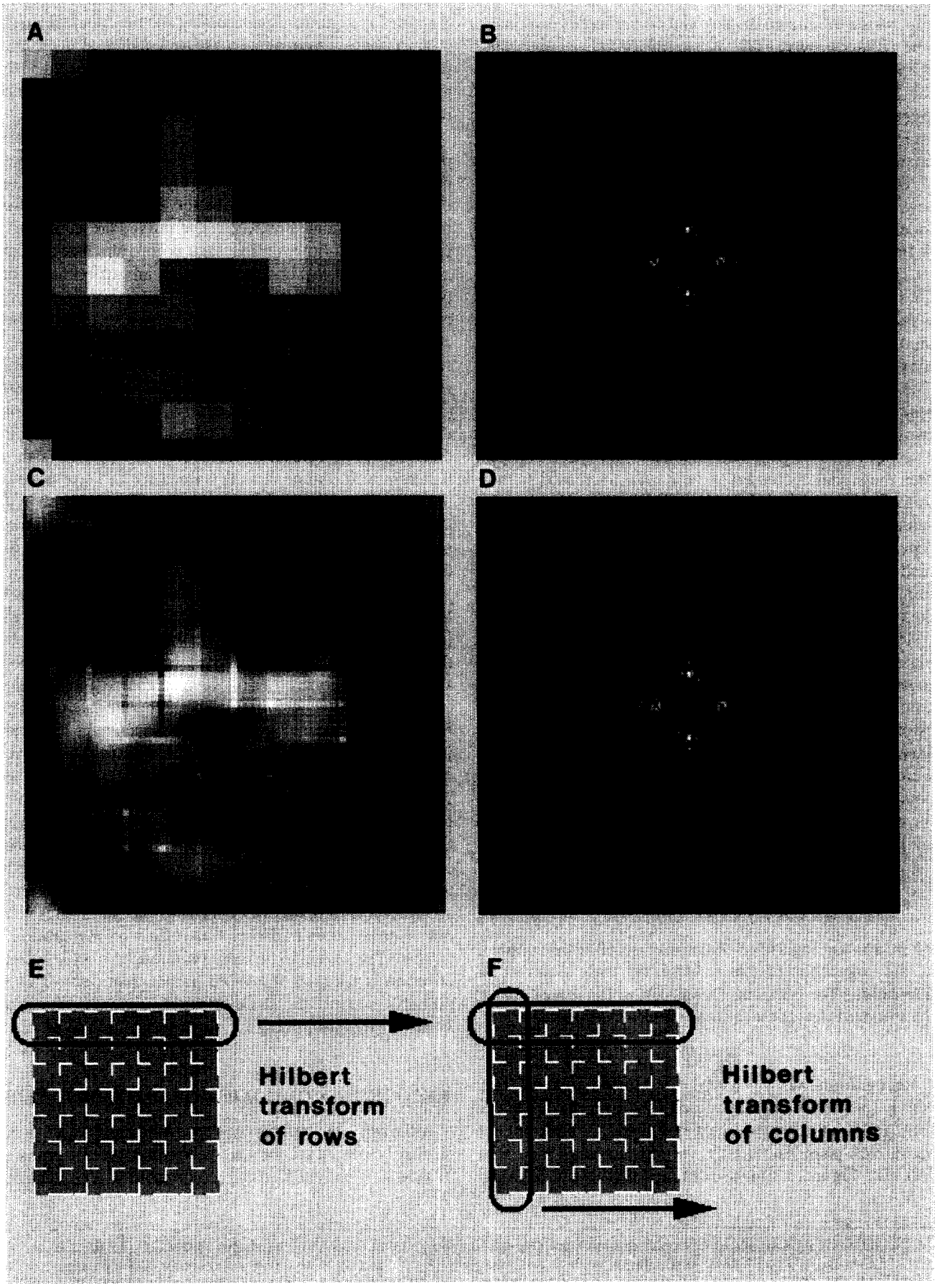
Noise corruption. As a performance index of recognizability, we chose to measure the maximum amount of noise that could be tolerated for reliable recognition (rather than simply per cent correct or reaction times). Signal and noise stimuli were treated separately, and summed on-line during the experiments. For each condition, 40 noise stimuli were prepared, each from an independent draw of white noise. The noise images were processed in the same way as the signal images (described above). They were filtered, rescaled to span the whole grey-scale range and blocked with the same regime as the faces or letters. Spurious harmonics, extracted by subtraction, were phase-shifted and resummed to the blurred image. During the experiment, signal and noise images were summed on-line, each with variable contrast. Although the signal and noise images were processed separately (to reduce the number of stored images), the final result is the same as would have been if the signal and noise were mixed before processing, as all the operations (integration, filtering etc.) were linear. Examples of noisy blocked stimuli are shown in Fig. 3.

All images were prepared in advance, and stored on disk. In practice, 120 different images were prepared for the faces (six faces in ten different poses, plus their mirror reflections) and 40 samples of noise. For all stimuli, there was the blurred condition, the standard blocked condition, and the blocked condition with the higher harmonics shifted (seven different phase shifts), yielding a total of 2880 images, each 128×128 pixels. Contrasts of the images are expressed as RMS contrast, the standard deviation of the images divided by the mean. The average RMS contrast on the screen was 0.15 for the faces, and 0.22 for the noise, at maximum.

Psychophysical procedures

Subjects were required to identify the face in a six-alternative, forced-choice paradigm, by pressing one of six response buttons. The images were presented on a Barco Calibrator monitor, driven by a framestore (Cambridge Research VSG) under computer control (IBM

FIGURE 2 *(facing page).* (A) and (C): A blocked image and one with the spurious components phase-shifted by $\pi/2$. (B) and (D): The power spectra of the high frequencies (>8.2 c/picture width) of (A) and (C), showing that it is unaffected by the phase manipulations. The low frequencies have been omitted to display the spurious frequencies more clearly. (E) and (F) illustrate the technique for phase-shifting the two-dimensional spectrum by $\pi/2$. The Hilbert transform was calculated for each row of the original spurious image by adding $\pi/2$ to the phase spectrum and multiplying the amplitude by sign function, then applying the inverse Fourier transform to produce an intermediate image. We then calculated the Hilbert transform for each column to produce the $\pi/2$ phase-shifted image, comprising the high-frequency components of (C).


A
B
C
D
E
F
Hilbert
transform
of rows
Hilbert
transform
of columns

compatible PC). The PC read the images from the Silicon Graphics disk (via PCNFS network software), and supervised the image display and collection of responses. The images were all 128×128 pixels, subtending 9×9 cm on the display screen, subtending 5 deg from the viewing distance of 1 m. Signal and noise images were interleaved on alternate frames (framerate 170 Hz), and the contrast of each varied independently by changing look-up tables (with appropriate linearization).

To minimize temporal transients that may favour the low spatial frequencies of the faces, the images were faded in and out of the noise gradually, as shown in Fig. 4. Before each trial, the contrast of the noise was maximal (0.2) and the signal attenuated by 2 log units to 0.0015. On initiation by the subject (time 0 in Fig. 4), the signal was faded in and the noise faded out exponentially, both at 0.1 log units every two frames. The signal reached maximum contrast (0.15) after 40 frames (230 msec). The noise continued to decrease only until it reached the appropriate level for that trial: Fig. 3 shows examples for two noise contrasts. The display remained constant for 1 sec after the signal had reached maximum contrast, then the signal decreased and the mask increased to their initial values. After the subject had responded, a new sample of noise was displayed, and the procedure repeated after a short interval. The new images were read from disk while the subject responded to the previous trial, which did not, in practice, cause noticeable delays.

The contrast of the noise (during the presentation plateau) varied from trial to trial, guided by the QUEST routine to home in near threshold. However, to ensure coverage of a range of noise contrasts, the QUEST estimate was randomly perturbed by ± 0.1 log units. At least five separate QUEST sessions, each of 40 trials, were run, and the final estimates of threshold were obtained by fitting a cumulative gaussian distribution to the probability of seeing against log noise contrast, by the simplex technique (Nelder & Mead, 1964):

$$f(s) = \gamma + \frac{1-\gamma}{\sigma\sqrt{2\pi}} \int e^{-0.5\left(\frac{s-s_0}{\sigma}\right)^2} ds \qquad (1)$$

where s is the logarithm of the inverse of noise contrast, s_0 the logarithm of the inverse of noise contrast at threshold, γ the guessing factor (1/6 in this case), and σ the standard deviation (in log units).

Results

Recognition was measured for blurred images [Fig. 1(A)], for the images quantized at 11 pixels per block [Fig. 1(B)], and for the blocked images with the higher harmonics phase-shifted by various amounts [see Fig. 1(D–F)]. Figure 3 shows examples of the images with 0.002 noise contrast (left-hand figures) and 0.046 noise contrast (right-hand figures). It is apparent that the noise degrades recognition, particularly for the blocked image in zero phase. During each experimental session, at least four conditions were run simultaneously, with the threshold noise contrast for each condition determined by independent QUEST staircases.

Figure 5 shows two examples of how recognition performance varied with noise contrast, for a blocked pattern and one phase-shifted by π. The percentage correct was systematically better for the phase-shifted pattern at all noise contrasts. The curves are cumulative gaussian fits, following Eq. (1). It is apparent that both the mean of the fit (s_0) and the slope (σ) are different for the two conditions. The dotted line corresponds to 58% performance, intersecting the curves at $s = s_0$. The dashed curve corresponds to the 74% performance level, half a standard deviation ($s = s_0 + \sigma/2$) higher. A probable reason for the shallower functions for the 0 phase shifts (supported by subjective reports), is that despite our efforts to eliminate local cues for recognition (20 different views of each face, with gross features such as hair removed), some cues remained for at least some of the faces, enabling subjects to guess the response with greater than chance performance. It is, therefore, probably more reasonable to consider the higher criterion of 74% performance.

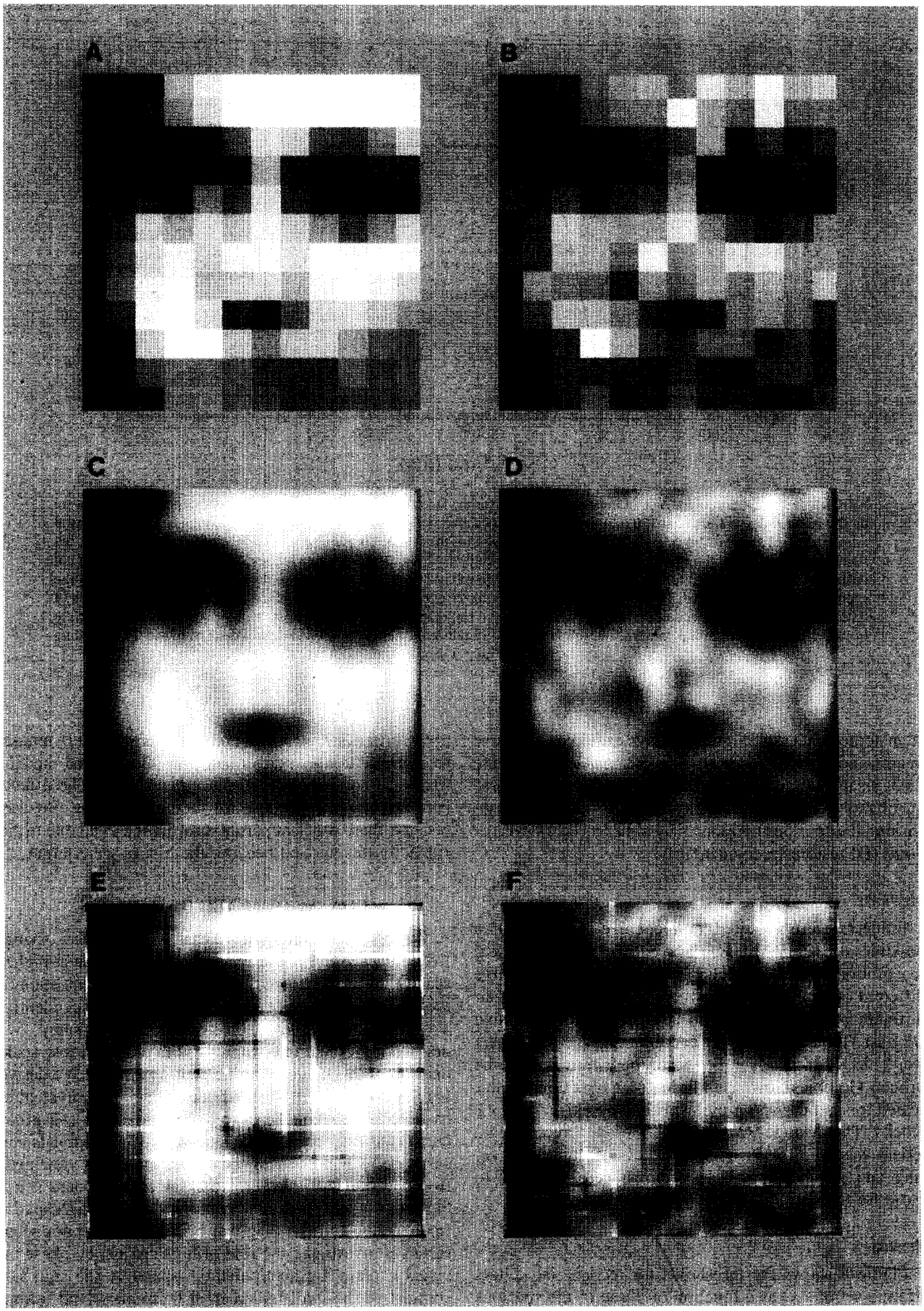
Noise contrast thresholds are shown in Fig. 6, for both the 58% criterion (filled symbols) and the 74% criterion (open symbols). The broken lines show performance for the blurred faces (dashed: 58%, dotted: 74%). For both subjects, for both criterion levels, coarse quantization clearly decreased performance well beyond that expected from the reduction of spatial frequency content (compare performance for 0 phase shift with that for the blurred images). The degree to which performance was affected varied with subject and criterion level, ranging from a factor of 2 to a factor of 20. However, when the phases of the higher harmonics were displaced, performance steadily recovered. For the maximum phase shifts of π, the results were very similar to those of the blurred stimuli, indicating that when the spurious harmonics were out-of-phase, they had very little effect on recognition.

The effects seem to be more pronounced if 74% correct criterion is taken as threshold, rather than 58%. This is probably because the slope of the psychometric function varies with phase, being far shallower at 0 than at π.

RECOGNITION OF COARSE QUANTIZED LETTERS

Although the effects of coarse quantization were originally described for face recognition, most types of images will be rendered unrecognizable when quantized

FIGURE 3 *(facing page)*. Examples of stimuli used for the face recognition experiment. The images on the left [(A), (C) and (E)] have only 0.002 RMS noise contrast, while those on the right [(B), (D) and (F)] have 0.046 noise contrast. The top couple [(A) and (B)] have spurious components in 0 phase, the middle couple [(C) and (D)] have no spurious harmonics and the bottom couple [(E) and (F)] have spurious harmonics in $\pi/2$ phase. Adding noise makes all images harder to recognize, particularly when the spurious components have not been phase-shifted (B).


A
B
C
D
E
F

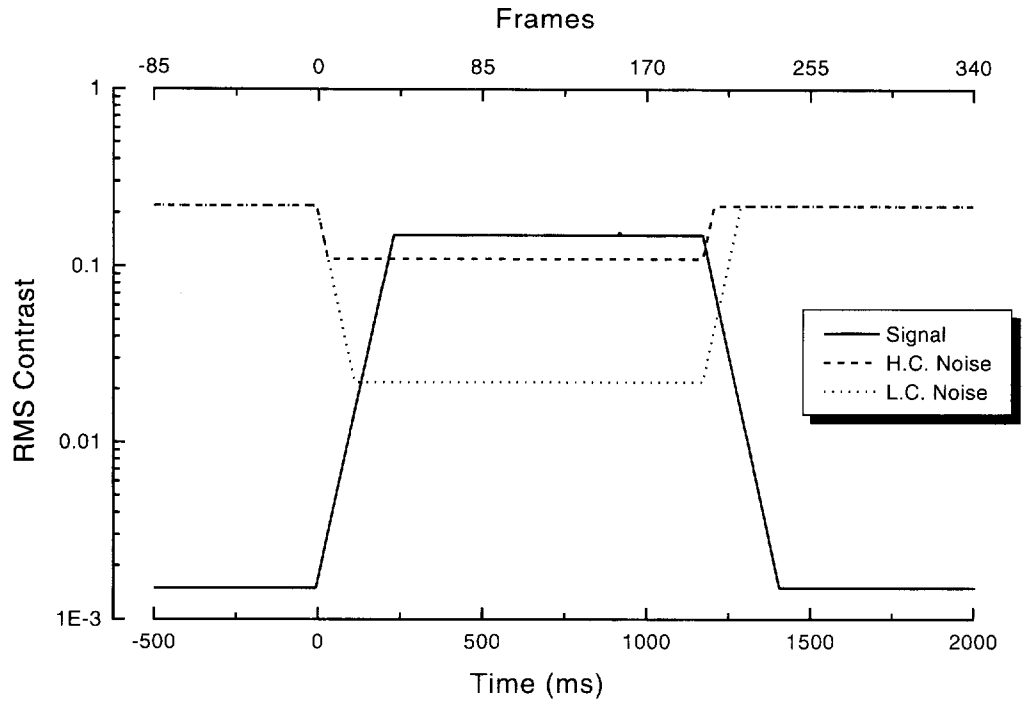


FIGURE 4. The temporal sequence of image presentation for the face and letter recognition experiment. Signal and noise were displayed on alternate frames, at 170 frames/sec. Before of each trial, the contrast of the noise was maximal (0.2) and the signal attenuated by 2 log units, to 0.0015 for the faces, and 0.004 for the letters. At the beginning of each trial (time=0 msec), the signal was faded in and the noise faded out exponentially, both at 0.1 log units every two frames. The signal reached maximum contrast (0.15 for faces, 0.3 for letters) after 40 frames (230 msec). The noise continued to decrease only until it reached the appropriate level for that trial. The dashed lines show an example for high contrast noise, 0.1, and the dotted line for a contrast of 0.02. The display remained constant for 1 sec after the signal had reached maximum contrast, then the signal decreased and the mask increased to their initial values. After the subject had responded, a new sample of noise was displayed, and the procedure repeated after a short pause.

at an appropriate rate. Figure 7 shows examples of coarse quantization of the letter R. The original letter is quite recognizable after heavy blurring, but not after coarse quantization to an 8×8 array (A). As before, phase-shifting the spurious harmonics by either π (B), $\pi/2$ (C) or $-\pi/2$ (D) restores recognizability.

Methods

Subjects were required to recognize letters from a possible six (C, D, K, N, R or S). To minimize local recognition cues, the letters were displayed at five different orientations (between ± 60 deg) and four different positions, and could be in positive or negative contrast (total of $6 \times 5 \times 4 \times 2 = 240$ images). The letters were 128×128 pixels, taken from the Times font of the Silicon Graphics, lowpass filtered (4.2 c/picture width), scaled to span the 256 grey levels and quantized to 8×8 blocks. The average RMS contrast on the screen for the letters was 0.3. The remaining details of image processing and psychophysical procedures were identical to those for the faces.

Results

Figure 8 shows the thresholds for recognizing letters (from a set of six) as a function of phase of the spurious harmonics. As with face recognition, performance (at 58% criterion) was severely impaired at 0 phase, recovering to the performance of the blurred images (dashed lines) at phase displacements of π. The effects here are larger than those for face recognition, possibly because there was less partial information available to aid subject guessing, and because the blocks were larger.

COARSE QUANTIZED SINUSOIDS

The results so far indicate that when the low spatial frequencies are in phase-congruence with those at higher scales they are "captured" by the grid like organization of the higher scales, and cannot be accessed independently. To study better the mechanisms of the capture, we devised a more simple stimulus with only two low-scale harmonics that can be manipulated separately. We derive our stimuli from the checkerboard, which is equivalent to two crossed diagonal sinusoids, quantized at $\sqrt{2}$-times the spatial frequency of the sinusoids. The diagonals can easily be seen by blurring the checkerboard, or viewing it from a distance. However, when the checkerboard is not blurred, the sinusoids are "captured" by the high spatial frequencies, and the diagonal orientations are not seen. Here we investigate the role of high-frequency phase on visual capture, by manipulating separately the sign and contrast of the two fundamental harmonics, and asking observers to indicate the predominant orientation. The

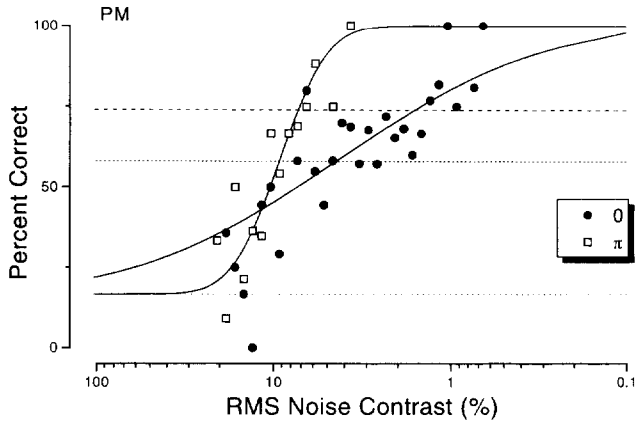


FIGURE 5. Two examples of how recognition performance varied with signal strength. Filled circles refer to images with spurious components unchanged [like Fig. 1(C)], and open squares to images with spurious components shifted by π [like Fig. 1(D)]. The abscissa plots the RMS contrast of random blocked noise. Performance was better for the π phase-shifted stimuli at all noise contrasts. The fitted curves are cumulative gaussians, described by Eq. (1), bottoming out at the guessing level (γ) of 17% (dotted line). There is a clear difference, not only in the mean of the fit, but also in the slope. The short-dashed line shows 58% performance ($s = s_0$) of Eq. (1), and the long-dashed line 74% performance, half a standard deviation higher ($s = s_0 + \sigma/2$).

rationale behind this approach is that if one of the components is captured by the higher frequencies, it will not contribute to the perceived orientation.

Methods

The original image was an 8×8 checkerboard of maximum contrast, of 128×128 pixels. Viewed from 1 m, the display subtended 5 deg (as before), so each square subtended 37.5 arc min. Figure 9 shows examples of a checkerboard [Fig. 9(A)] and the stimuli derived from it. Figure 9(B) shows a highpass filtered checkerboard (with the fundamentals removed), similar to the "spurious harmonics" of the blocked faces or letters. Figure 9(C) shows the two fundamental harmonics. However, in this image, the +45 deg (right tilting) fundamental has been contrast-reversed, to probe the effects of high-frequency phase on capture. Figure 9(D–F) were constructed by adding to the low harmonics of Fig. 9(C) the higher spurious harmonics, phase-shifted by either 0, $\pi/2$ or $-\pi/2$, respectively.

To quantify the effects of phase on capture, we varied the relative contrast of the two fundamentals to seek a null point in the perceived orientation. One fundamental sinusoid was left unaltered and the other was contrast-reversed (phase-shifted by π): Fig. 9(C). The higher harmonics were also left at their original contrast, but their phase was varied over a range of $\pm\pi$, before resumming to the low-frequency pattern. In practice, the pattern was "flipped" at random from trial to trial (inverted symmetrically around the vertical midline), so either the unaltered or the contrast-reversed fundamental could be oriented to the left or right. The subject's task was to indicate whether the dominant tilt was ± 45 deg. The contrast of the contrast-reversed fundamental varied depending on response, tending towards the contrast to support ambiguous apparent orientation (guided by the QUEST routine, with ± 2 dB of randomization). At least five different conditions were randomly interleaved

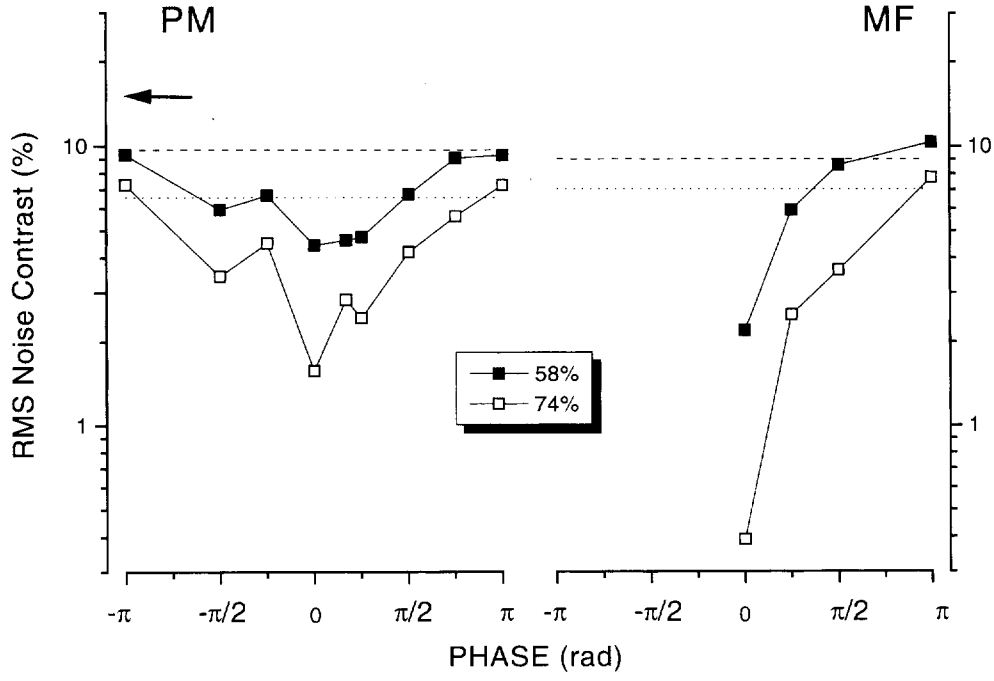
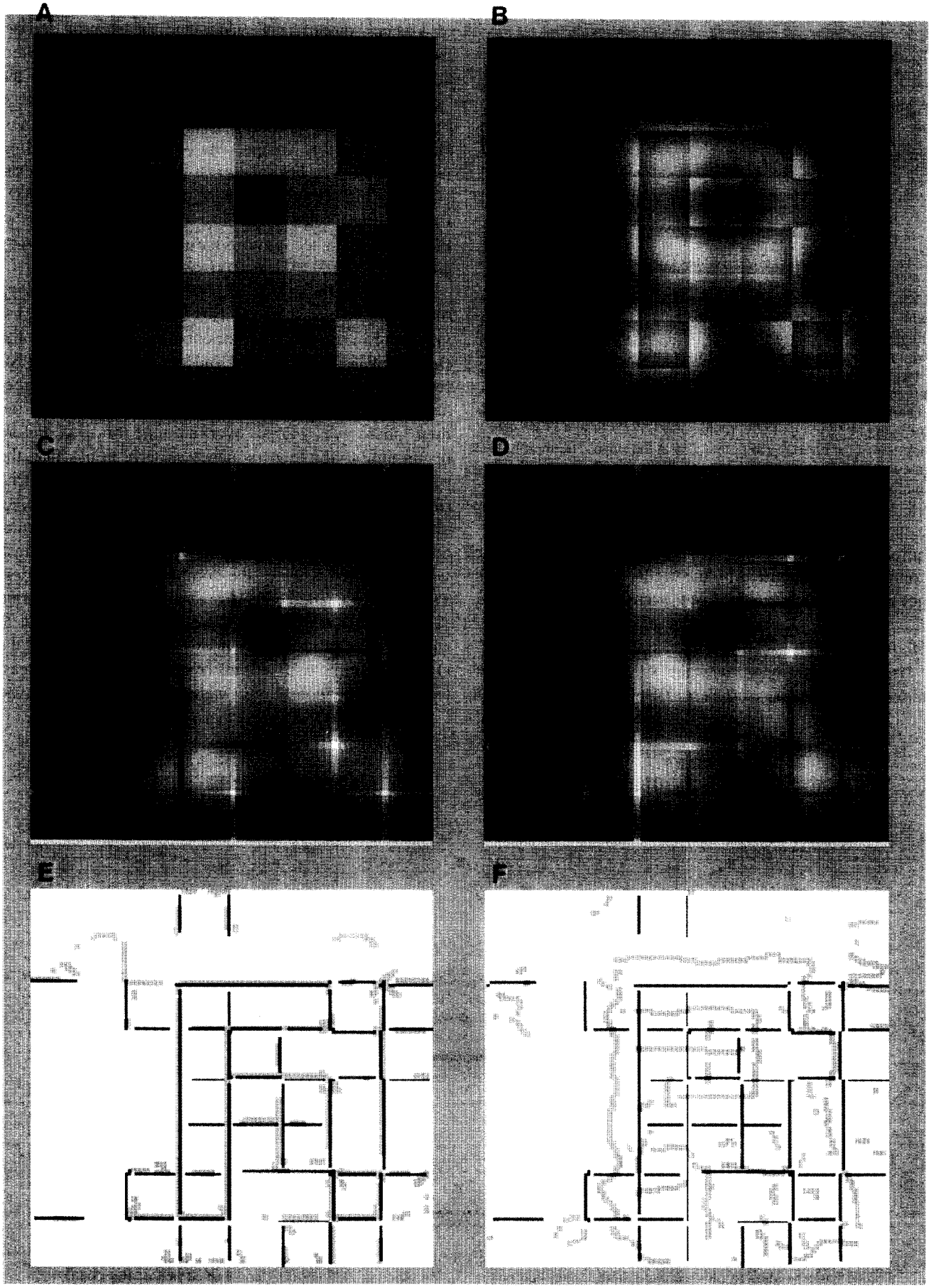


FIGURE 6. Recognition performance for two naïve observers for various phase shifts. Two levels of sensitivity are reported, one corresponding to 58% correct (short-dashed line of Fig. 5), the other to 74% correct (long-dashed line of Fig. 5). The ordinate shows the contrast of the noise at threshold, with the arrow indicating the average RMS contrast of the blocked faces. The dashed and dotted lines show 58% and 74% sensitivity thresholds to blurred faces.


A
B
C
D
E
F

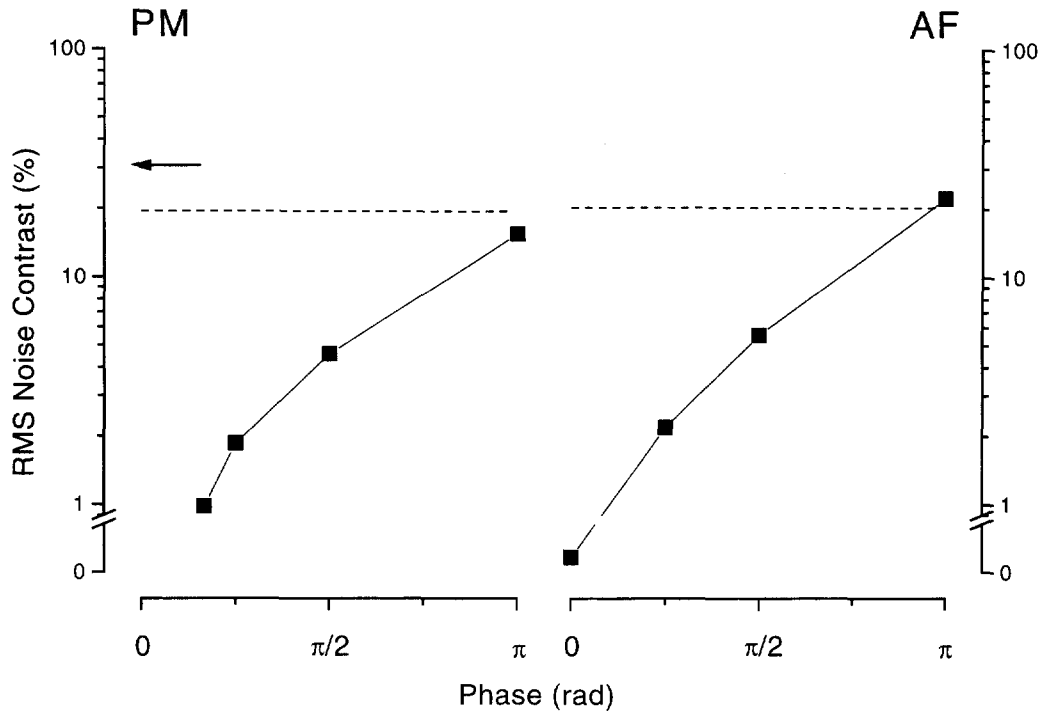


FIGURE 8. Threshold performance for letter recognition for two naïve observers, for various phase shifts. Threshold was taken as the contrast of noise to yield 58% correct responses. The arrow indicates the average RMS contrast of the blocked letters. The dashed lines show the sensitivities for the blurred letters, with no spurious components.

during a single session, minimizing response stereotyping.

About 200 trials were measured for each condition, over four or five experimental sessions. Again the results were fitted with Eq. (1), with s describing log-contrast and guessing rate $\gamma = 0$. For the contrast sensitivity measurements of the masking studies, the test (sinewave) was interleaved with the mask so its contrast (reported as Michelson contrast) could be controlled on-line. The results were also fit with Eq. (1), with $\gamma = 0.5$ (two-alternative, forced-choice).

Results

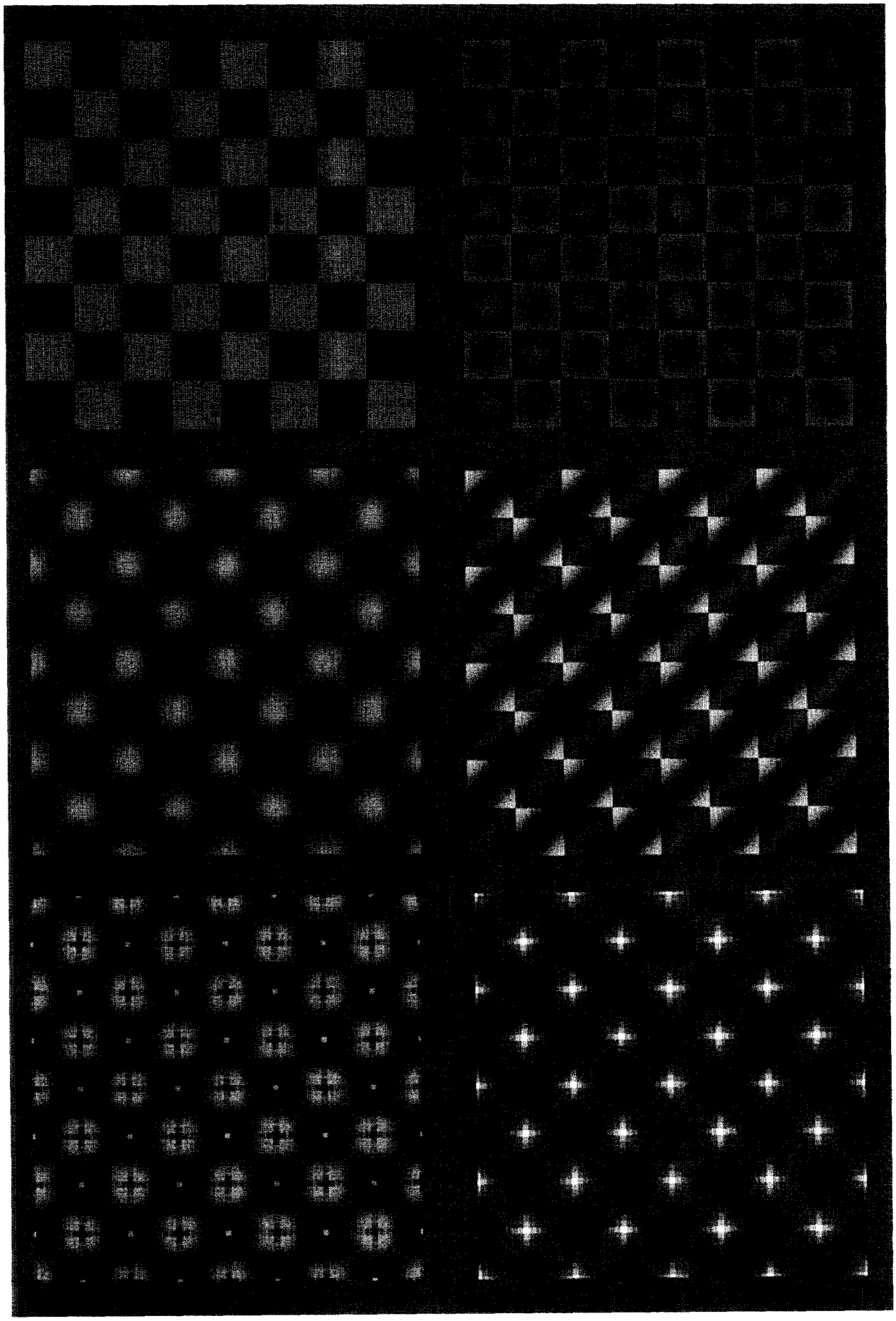
The effects of the phase of the spurious frequencies on apparent orientation are clearly illustrated by inspection of Fig. 9. The standard checkerboard appears as a vertically–horizontally organized structure, although there is actually no energy along those orientations. When one of the fundamentals is reversed in contrast (leaving the higher harmonics at phase 0), it is seen clearly in transparency against the rest of the pattern, and the orientation of this pattern seems to dominate the tilt of the stimulus. Now, when the phase of the higher harmonics is changed, for example by $\pi/2$ as shown in Fig. 9(E), both are seen in transparency. When the phase of the higher harmonics is shifted by $-\pi/2$ [Fig. 9(F)], so all harmonics are in peaks-add phase, the orientations are again balanced, but without transparency.

The effects of changing the phases of the higher harmonics can be illustrated quite dramatically by photographing or photocopying Fig. 9(B) and Fig. 9(C) onto separate transparencies, and superimposing them on an overhead projector. Slowly sliding one over the other changes the phase coherence from one fundamental to another, changing the apparent orientation of the pattern. Note that if the patterns are moved too quickly, the orientations no longer appear to vary, indicating that capture takes some time.

Figure 10 shows the percentage of trials where the perceived orientation was given by that of the reversed-contrast fundamental, as a function of its contrast, for five different conditions of higher harmonic phase. The condition with the blurred stimuli (no higher harmonics) is shown by the thick curve (without data points). This serves as a control and gives an idea of the precision of the subjects at this task. When the contrasts of the two sinusoids are matched (dashed vertical line), subjects report the orientation of the reversed-contrast fundamental on 50% of the trials. As the contrast of this fundamental increases, so does the proportion of trials where it dominates the perceived orientation.

When the higher harmonics are phase-shifted by $\pm\pi/2$

FIGURE 7 *(opposite)*. (A) Example of a letter "R", coarse quantized into 8×8 blocks. In this example the letter is vertical, but others were rotated by ± 15 or ± 30 deg. The average RMS contrast was 0.4. (B), (C) and (D): The blocked letter with the spurious harmonics phase-shifted by π, $+\pi/2$ and $-\pi/2$, respectively. (E) and (F) show local energy simulations of (A) and (B). The thin black lines are the peaks in local energy at 24 c/picture width, and the thick grey lines at 6 c/picture width (see section on *Simulation of results*.



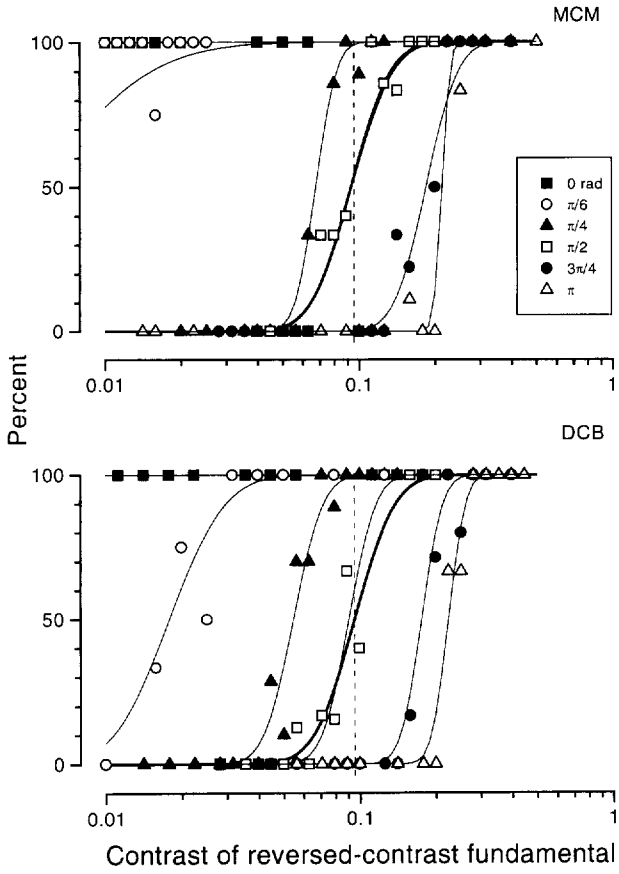


FIGURE 10. Per cent of trials on which the orientation of the pattern was reported as being the same as that of the reversed-contrast fundamental, for various phase shifts of the higher components [Fig. 8(D)–(F) show examples of the stimuli]. The sigmoid curves are best-fitting cumulative gaussians, following Eq. (1) with $\gamma = 0$. The thick sigmoid is the fit to the data of apparent orientation of the lowpass pattern [Fig. 8(C): data points not shown]. The vertical dashed line indicates the Michelson contrast of the unaltered fundamental, so at this point the contrasts of the two fundamentals were balanced. As may be expected, the averages of the curves with no higher frequencies pass through this point, as do the curves for the $\pi/2$ phase-shifted higher harmonics. The remaining curves are systematically displaced along the abscissa, depending on the phase shift of the higher components. At 0 phase shift, the contrast-reversed fundamental dominated at all contrasts, even when it was physically absent (in agreement with Burr *et al.*, 1986).

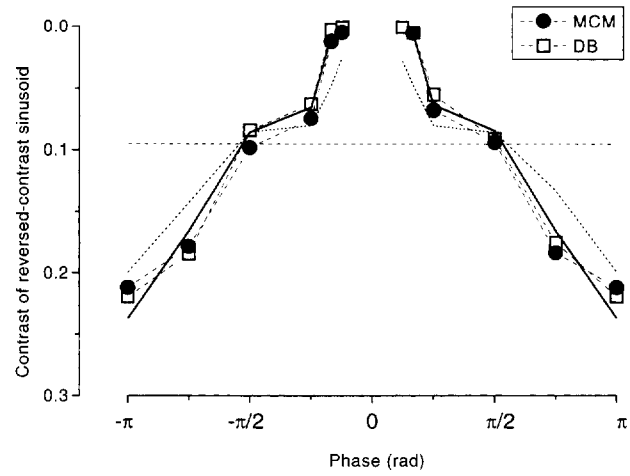


FIGURE 11. Contrast of the contrast-reversed fundamental at which the apparent orientation was balanced (the 50% point of the fitted curves of Fig. 9), as a function of the phase shift of the higher components. The horizontal dashed line shows the contrast of the unaltered fundamental. The thick continuous curve shows the prediction from the local energy model, as described in Fig. 14, for a threshold *orientation bias* of 2, and the short dashed line for an orientation bias of 1.4.

(falling between the phases of the two fundamentals), the results are similar to that of the blurred stimuli, both in null point and width of psychometric function (the curves of MCM are actually superimposed). This shows that adding high-frequency components to the pattern imposes no extra difficulty for the subjects to determine prevailing orientation. Varying the phase shift of the higher harmonics systematically moves the curves along the contrast axis. For phase shifts near 0 (filled squares), the phases of the higher harmonics are similar to that of the unaltered fundamental, so it is "captured", leaving only the reversed-contrast fundamental to dominate the perceived orientation at all contrasts. Indeed, at phase 0, the null is not possible at any contrast: even when the contrast is zero, the apparent orientation is given by this fundamental that is not physically present (see also Burr *et al.*, 1986). For phase shifts greater than $\pi/2$, the phases of the higher harmonics are more similar to that of the reversed-contrast fundamental, capturing it. The unaltered fundamental now dominates the apparent orientation, so the contrast of the other fundamental needs to be raised to produce a balance in orientation.

Figure 11 plots the amount of contrast of the reversed-contrast fundamental required for the orientation null (50% points of psychometric functions such as those of Fig. 10) as a function of phase. The results of both observers are very similar, showing the clear, almost linear relationship between contrast and phase.

Masking

The original explanation for the blocking effect was "critical band masking" of the low spatial frequencies carrying the image information by the higher spurious frequencies. The fact that the effects depend so critically on phase cast some doubt on this assertion, but the possibility remains that the masking effects themselves are phase specific. We tested this idea by measuring

FIGURE 9 *(opposite)*. Examples of various images derived from the checkerboard. (A) An 8 × 8 checkerboard. (B) A highpass filtered checkerboard (ideal, above 5.7 c/picture). (C) A lowpass filtered checkerboard, with the 45 deg fundamental reversed in contrast (phase-shifted by π). (D) The unfiltered checkerboard with the 45 deg fundamental reversed in contrast (equivalent to the sum of (B) + (C). (E) and (F) Same as (D), but with the higher harmonics phase shifted by $+\pi/2$ and $-\pi/2$, respectively. If (B) and (C) are photocopied onto separate transparencies and superimposed on an overhead projector, the prevailing orientation of the pattern can be changed from 45 to −45 deg by sliding image (B) by one block width (half a cycle of fundamental).

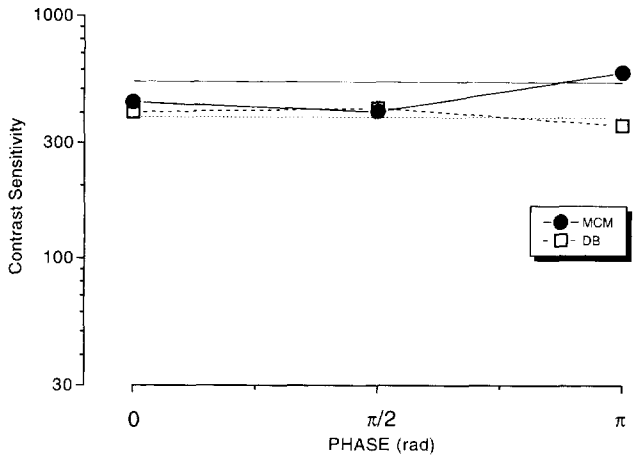


FIGURE 12. Contrast sensitivity for discriminating the orientation of a single fundamental of a checkerboard, for various phase shifts of the higher components (the orthogonal fundamental was removed). The continuous and dashed horizontal lines show the thresholds in the absence of any higher components for MCM and DB, respectively.

detectability of a sinusoid, in similar conditions to those used for the apparent orientation judgements. In this experiment, the test was a single sinusoid, oriented ± 45 deg, either on its own or in the presence of the higher harmonics, phase-shifted as described in the previous section. The subjects were required to detect the orientation of the test, in a two-alternative, forced-choice procedure. Contrasts were again guided by the QUEST procedure, and thresholds obtained by fitting gaussians to the probability of seeing curves [Eq. (1) with $\gamma = 0.5$].

The results are shown in Fig. 12. It is clear that at all phases, the higher harmonics had very little effect on contrast sensitivity. The sensitivities for all phases were very similar to those measured without the higher harmonics. Thus, the higher harmonics affect the apparent orientation of the pattern, not by "masking" the visibility of the lower harmonics, but by more subtle processes.

SIMULATION BY THE LOCAL ENERGY MODEL OF THE TRANSPARENCY EFFECTS

In this study we have shown that changing the phase of the spurious high spatial frequencies' harmonics greatly reduces, or destroys completely, the blocking illusion: after a large phase shift, the stimuli are perceived as two independent structures in transparency, with no "capture" of the low frequencies. In this study, the phases of the entire spectrum of the spurious frequencies were displaced together, so the relationships between the phases of various harmonics remained unaffected. The location where all the phases were most similar was therefore unchanged (see Morrone & Burr, 1988), so the energy functions at high-scales were the same at all phases. Therefore, the localization by the local energy model of features at high-scales is not affected by this transformation. However, the energy map of the medium and low-scales that respond to both the signal and the spurious frequencies will be greatly affected by the phase manipulations, as local energy is maximum at points of the image where the arrival phases of various components within its bandwidth are most similar (see Morrone & Burr, 1988). Changing the phase of some of the components breaks the phase congruence, changing greatly the local energy map.

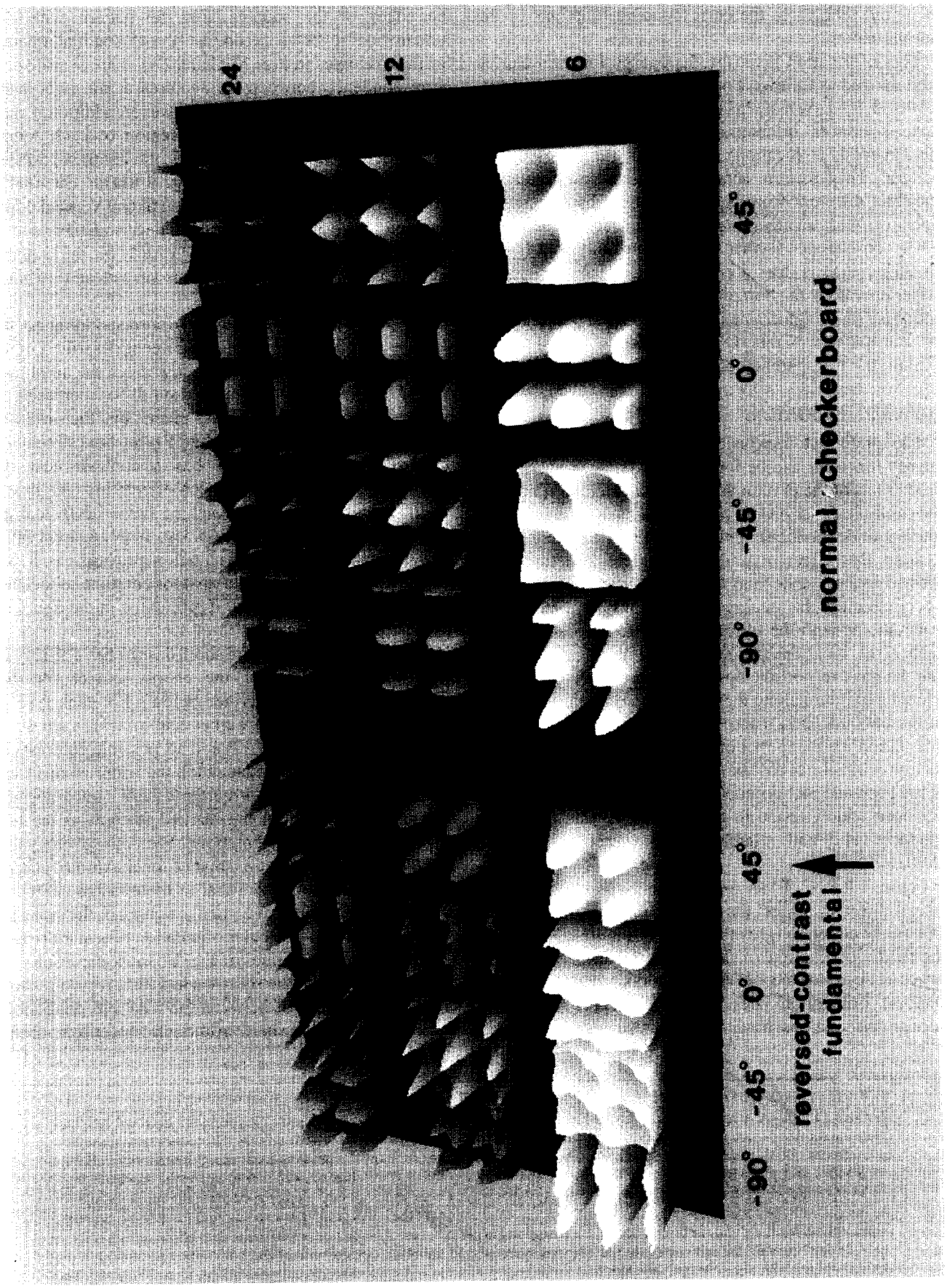
Figure 13 shows in three dimensions the energy functions obtained for different size and orientations of the linear operators, for one period of the normal checkerboard pattern and for the pattern with the +45 deg fundamental reversed in contrast (marked by the arrow in Fig. 13). The energy functions have been computed as the square-root of the sum of the square of the response of the even and odd filters (filter equation in caption). The filters are broadband both in orientation and size of about ± 45 deg and ± 0.28 log units, respectively at half-height.

For the checkerboard pattern, energy functions at all scales have clear maxima along the horizontal and vertical edges. The ridges are sharper for the high than the low-scale, but all maxima are spatially coincident. However, for the pattern with the contrast-reversed fundamental, there is no such correspondence between scales. The energy functions at all orientations at the highest scale, and at all the scales of -45 deg orientation, are very similar to those of the normal checkerboard pattern. However, all the remaining functions differ markedly: the +45 deg energy functions (at 12 and 6 c/picture width) have clear peaks corresponding to the centre rather than the edges of each square, almost the inverse of the pattern of energy at the orthogonal

FIGURE 13 *(facing page)*. The local energy derived from operators of various scales and orientations illustrated for a small section (four squares) of normal checkerboard (right) and checkerboard with the 45 deg contrast-reversed fundamental (left), computed as described in the text. The form of the even and odd linear filters (G_E and G_O)is given by:

$$G_E(f_x,f_y) = \exp\left[-\left(\frac{(\ln f_x/f_p)^2}{2\sigma_x^2} + \frac{f_y^2}{2\sigma_y^2}\right)\right],$$

where f_x and f_y are the frequencies orthogonal and parallel to the orientation of the operator, f_p is the peak spatial frequency and σ_x and σ_y determine bandwidth along the axes of the operator, set to produce a spatial frequency bandwidth of ± 0.28 log units and an orientation bandwidth of ± 45 deg. We used 16 operators, spanning four spatial scales (3, 6, 12 and 24 c/picture width, corresponding to 0.75, 1.5, 3 and 6 c/horizontal-period) and four orientations (0, ± 45 and 90 deg). With the unaltered fundamental, the peaks in local energy follow the outline of the checks, at all scales and orientations. When the +45 fundamental was contrast-reversed, the energy maps for the low and medium scales at 45 deg change considerably.


24
12
6
-90°
-45°
0°
45°
-90°
-45°
0°
45°
reversed-contrast
fundamental
normal checkerboard

orientation. At 6 c/picture, the horizontal and vertical orientations show ridges that are dislocated by a quarter of period relative to the corresponding functions of the normal pattern.

The information given by the energy function can be summarized as follows. For the pattern with the contrast-reversed fundamental, maximum energy is produced by large operators oriented along the direction of this fundamental, at locations away from the edges of the pattern. This pattern should dictate a low-frequency structure different from the one imposed by the smaller operator, with a prevailing orientation at +45 deg. The other consequence is that the structure imposed by the smaller operators does not have corresponding energy at larger operators: the horizontal and vertical border should comprise only high-frequencies and should be perceived as transparent "Craik–O'Brien" edges. This would explain the transparency effects that we see.

Quantitative predictions

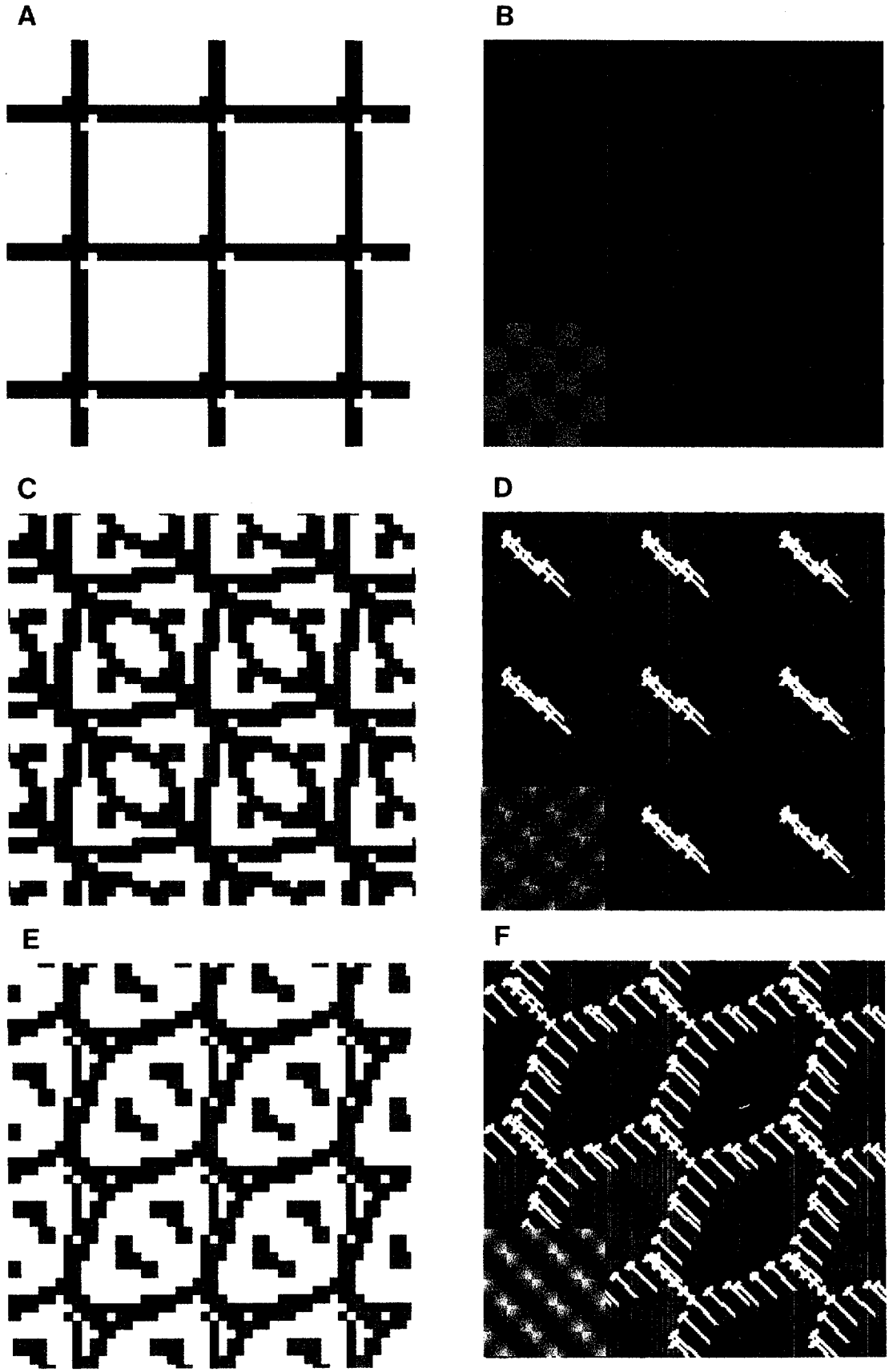
The qualitative conclusions reached by inspecting the energy functions can be formalized to make quantitative predictions of the psychophysical data, after defining how the multiple functions at various scales and orientations interact to define the spatial structure of the stimulus. We assume that separate and independent structures are defined at each scale, and that these structures are given by combining the energy distributions at the various orientations, with the following two-step procedure. For each pixel, local energy is evaluated at all orientations, and the most responsive operator is chosen. If the point is a local maximum along the orientation orthogonal to that of the operator, it is marked as an oriented feature, with orientation matching that of the operator. The type of the feature (edge–line) will be given by the relative strength of the even and odd symmetric linear filters. This strategy has been widely tested in other situations and shown to produce valid and reliable results (Perona & Malik, 1990).

Figure 14(A, C, E) shows the energy maxima marked at two different scales: the 24 c/picture (black thin line) and the 12 c/picture (grey thick lines). Figure 14(B, D, F) shows the orientation of the features marked at the 12 c/picture scale for the +45 (red vectors) and −45 (yellow vectors) orientation. For clarity, the horizontal and vertical orientations have been represented with black points. For the normal checkerboard [Fig. 14(A, B)] the peaks at the two scales overlap in space and coincide with the edges of the squares. For the pattern with one fundamental reversed in contrast [the 45 deg fundamental for the examples in Fig. 14(C and D)], there is still some correspondence between the features at the two scales along the edges of the squares. However, additional clear features emerge in the internal region of each square corresponding to the orientation of the contrast-reversed fundamental [red vectors of Fig. 14(D)]. The model would predict a structure of horizontal and vertical edges enclosing bars oriented along 45 deg, together with a −45 deg structure only at the vertex of the squares (yellow vectors). This prediction is very close to what is actually perceived.

For the pattern in Fig. 14(D) (corresponding to the stimulus labelled as π in the experiment described in the previous section), we measured how much the amplitude of the −45 deg fundamental should be increased to balance the orientation bias given by the +45 deg oriented features. Figure 14(E and F) show the simulation obtained for a pattern that has a contrast near the psychometric setting for balanced orientation (see inset). No features at scale 12 c/picture have horizontal or vertical energy, but only ± 45 deg. It is apparent [Fig. 14(F)] that the −45 deg orientation prevails, as it has more features belonging to that orientation associated with points of higher energy (represented in the figure by the length of the vectors).

The thick continuous, and short-dashed lines of Fig. 11 show the predictions of the energy model for two different thresholds (described below). For each type of stimulus (phase shift of higher harmonics), each scale below 24 c/picture (3, 6, 12) and all possible contrasts of the contrast-reversed fundamental, the feature maps were calculated to yield the "*orientation bias*", reflecting the ratio of local energy of the marked features at the two principal orthogonal orientations. We defined this value as the average of two ratios, that of the maximum energy at the two orientations and that of the average energy at the two orientations. We then calculated the contrast value to produce a particular orientation bias at all scales. The thick continuous lines of Fig. 11 show the results for a threshold of 2, and the short-dashed lines for a threshold of 1.4. The agreement between the data and predictions are good for both values. Although the choice of the

FIGURE 14 *(facing page)*. Illustration of how the apparent orientation thresholds were simulated from the local energy model for three patterns: a simple checkerboard [(A) and (B)], a checkerboard with the +45 deg contrast-reversed diagonal [(C) and (D)], and for this pattern with the unaltered fundamental at −45 deg of higher contrast [(E) and (F)]. The patterns are illustrated by the insets of (B), (D) and (F), respectively. The simulations on the left show the peaks in local energy functions of all four orientations, illustrated by the thick grey lines for the 12 c/picture width (3 c/horizontal-period) scale, and by the thin black lines for 24 c/picture width (6 c/horizontal-period). For (A) the energy at both scales all falls along the edges of the blocks, while for the images with the contrast-reversed fundamental much of the energy of the lower scale falls in between. The right-hand figures indicate the magnitude and the orientation of the energy at the peaks for the 12 c/picture width scale: the yellow vectors refer to the −45 deg operator, the red vectors to the +45 deg operator, and the black dots to vertical and horizontal operators. For the oblique orientations, vector length indicates magnitude (but not for the vertical and horizontal orientations, to avoid clutter). The orientation bias was calculated as the average of two ratios: the ratio of the maximum vector of each diagonal, and the ratio of the average vector length of each diagonal. For (D) the +45 deg vectors clearly dominate, while for (F) the −45 deg orientation prevails by a factor of two.


A
B
C
D
E
F

threshold orientation bias is somewhat arbitrary, it is important to note that this is the only free parameter required. Varying the criterion over a modest range had very little effect, except for phase shifts near $\pm\pi$. At phase 0, the simulation predicts an orientation bias of 0 in the absence of the reversed-contrast sinusoid, indicating that no features were oriented along the unaltered diagonal, so the missing diagonal always dominates perceived orientation.

These simulations were obtained using filters of ± 45 deg bandwidth in orientation and ± 0.28 bandwidth in spatial frequency. However, the pattern of simulation results was not peculiar to the choice of the filters, as similar results were obtained with different spatial frequency bandwidth (from ± 0.2 to 0.42 log units) with the same length-to-width aspect ratio. However, for narrower bandwidths more operators were necessary to span the different scales and orientations of the image spectra. Decreasing the width to length aspect ratio also increased the number of oriented filters necessary for accurate feature localization (see Perona & Malik, 1990).

Simulation of blocked letters

The same model was used to generate the feature maps of the quantized letter of Fig. 7. Figure 7(E) shows the simulation for the stimulus at phase 0, while Fig. 7(F) shows that for phase π. The features marked at the higher scale (24 c/picture) are aligned along the horizontal and vertical edges of the original blocks for both stimuli. However, features of the middle scale (6 c/picture) are aligned along the block borders only for in-phase stimulus. For the out-of-phase stimulus, the features at the two scales are quite separated. Again, the model makes a clear qualitative prediction that the out-of-phase stimulus should be seen as a letter "R" simultaneously superimposed on a grid structure.

It is interesting that the superposition should give rise to a transparency effect. The edge structure is present only at high spatial frequencies, so each block should be perceived as Craik–O'Brien edges. With normal course quantized images, the two feature maps correspond in space, so each high-frequency edge is in direct correspondence to the low-frequency one, so they should be perceived as solid and opaque. This leads to a suggestion about the nature of transparency. If we assume that separate feature maps are built at each scale, then the perception of transparency or solidity could depend entirely on whether there is correspondence between the feature maps at each scale. We will take this point up again in the Discussion.

DISCUSSION

Effects of blocking on recognizability

Although the blocking illusion of Harmon and Julesz is one of the strongest visual illusions, it has proven difficult to obtain quantitative measures of the strength consistent with the qualitative magnitude of the effect. The main reason for this probably arises from experimental difficulties. For small images, blocking certainly reduces recognizability beyond that expected by blurring out the higher frequencies (Uttal *et al.*, 1996a), but as the spurious frequencies are near the resolution limit, the effects would be expected to be smaller than for larger images. The use of larger images overcomes this problem, but creates others. The information about the face will be at very low absolute spatial frequencies (in cycles per degree). Brief exposure (customary for psychophysical measurements) greatly enhances sensitivity for low spatial frequencies at the expense of high frequencies (eg. Burr, 1981), boosting the image information. Furthermore, abrupt removal of a low-frequency signal leaves a vivid negative afterimage for some time after (e.g., Burbeck, 1986), that could be used for recognition (especially for letters, where the dark letters are as recognizable as bright letters).

We attempted to control for these problems by using quite large images, and not presenting them abruptly, but fading them in and out gradually. More importantly, they were gradually exchanged with random noise during both the rise and fall of presentation, which should have helped to mask the afterimage left by the low-frequency image. During the development of this technique, we found that recognition of blocked images was very easy, unless we took these efforts to eliminate artefacts.

Another difficulty with this class of experiment is that for a finite set of images, images may be recognized on the basis of some simple learned cue, such as a particularly white block in a certain position, etc. Indeed, discrimination of some coarse quantized images is particularly simple (Uttal *et al.*, 1996a). To minimize this possibility, we created a very large image set of both letters and faces, hoping to disrupt as many local cues as possible. However, particularly for the face recognition, subjects sometimes reported being able to make some discriminations on the basis of local cues. For this reason we suspect that the 74% recognition criterion is more valid than the 58% criterion. Indeed, this is why no results are shown for face recognition for the authors, as during development of the experimental techniques, they became over-familiar with the local cues in the image set. But in any event, it should be borne in mind that our data may underestimate the effects of blocking for all these reasons.

The sample psychometric curves of Fig. 5 show the importance of measuring thresholds, rather than simply per cent correct in a given condition. Without the addition of noise, performance approached 100% correct for both types of stimuli, under these conditions (with long exposure and unlimited response time). One may be led to conclude that blocking had no effect there. However, once the task was made more difficult by adding noise, there was a clear difference in performance. However, more importantly, this technique gives a quantitative measure of the magnitude of the effects, that cannot be calculated from simple per cent correct measures at a given difficulty level. Other strategies of image degradation, like those used by Uttal *et al.* (1996a, b) may also be

useful here, but it is important that the degradation used as the measuring parameter should not interact with the effect of blocking.

The results show not only that blocking reduces recognizability considerably, but that this effect was affected drastically by phase shifts of the spurious frequencies. This is obvious both from inspection of Figs 1, 2 and 3, and from the quantitative measurements. The effects of phase shift were at least an order-of-magnitude (sometimes greater), and phase shifts of π brought recognizability to levels near those of the blurred images, both for letters and for faces. It is important to realize that a phase shift of π is not like "phase-scrambling", in that it does not alter the distribution of high-frequency information over the image. The edges of the blocks remain where they were, of the same magnitude (but opposite sign) of contrast. The $\pi/2$ shift also does not affect the local RMS contrast (from Parseval's theorem), although it does increase the peak-to-peak Michelson contrast.

These results do not fit with the "critical band masking" theory of Harmon & Julesz (1973), as simple masking effects are not greatly affected by phase (e.g., Campbell & Kulikowski, 1966). Indeed, under the conditions of this experiment, we show that the presence of the higher harmonics of the coarse quantized sinusoid has virtually no effect on the detectability of a simple sinusoidal pattern, at any phase of the higher harmonics. We can, therefore, safely conclude that the vast impairment of visibility of the faces and letters does not result from simple critical-band masking, in which the low frequencies are rendered invisible to the visual system, but by more subtle means, discussed below.

The checkerboard as a coarse quantized image

The checkerboard has long provoked interest for vision research, largely because of its vertical–horizontal appearance, despite most of the Fourier energy being along the diagonals (eg De Valois *et al.*, 1979). Our previous work showed that removal of one of the diagonals altogether, creates a pattern with a strong sense of orientation along the orientation that contains no Fourier energy (Burr *et al.*, 1986). In that study we showed that the conditions under which the "missing" orientation dominates parallel closely those for the detectability of the fifth harmonic, suggesting that there need be at least two visible spurious harmonics for the illusion. The idea of local energy was not developed at the time of that study, but we would now say that the minimum of two higher harmonics was necessary to produce a peak in local energy to create the feature that structures the image into blocks, against which the "missing" fundamental is seen in transparency. Interestingly, this spatial phenomenon has a strong temporal analogy, the so called "non-Fourier motion" (Chubb & Sperling, 1988), where motion can be seen in a direction where there is no power in the Fourier spectrum. Both phenomena elicit a strong sense of transparency (see Fleet & Langley, 1994).

In the present study we have used a similar pattern to measure the effects of the phase of the spurious components, by asking observers to assess which diagonal predominates the perceptual appearance. The main advantage of this technique is that it is possible to collect large amounts of clean data for quantitative simulation. Although the observations are to some extent subjective, there was very good agreement of the results between subjects, and from one session to another (see Figs 10 and 11). We have reported detailed data only for two observers, but have also measured the effects on many colleagues, initially naïve as to the goals of the research, and all giving similar results. Most readers will agree from inspection of Fig. 9 that the effects of phase on apparent orientation are quite impressive.

Although the checkerboard can be considered to be a coarse quantized image, it is different from standard quantized images in at least one important respect: all of the "signal" is contained within two single harmonics of the same spatial frequency, about an octave lower than the lowest spurious frequency. This facilitates quantitative modelling, as there is no uncertainty about which frequency band contains the image information. One the other hand, with more complex blocked images, it can not be assumed a priori which band of frequencies contains most information, nor will the influence of the spurious components be the same for all spatial frequencies of the signal.

Simulation of results by local energy model

For the reasons mentioned above, it was possible to obtain accurate predictions for the appearance of the phase-manipulated checkerboard. With only one free parameter, we were able to predict with considerable accuracy the psychophysical measurements of apparent orientation. The procedure was simply to search for oriented maxima at all scales, and vary the contrast of the contrast-reversed fundamental until a threshold was reached.

With more natural blocked images, such as faces and letters, the situation is more complicated, and therefore more difficult to model quantitatively. However, the same principles probably hold. Figure 7(E and F) show the feature maps at high and medium scales, for the original and phase-shifted images. For the phase-shifted image, the medium scale more or less describes the letter "R", while for the original blocked image the feature maps at this scale follow the outlines of the blocks. At this scale, the spurious components interact with the components defining the letter and, providing that the phases are not too different, the spurious harmonics will dominate the feature map. Shifting the phase of the spurious harmonics breaks the phase coherence within this scale, allowing the peaks to be dictated by the low spatial frequencies. The fact that the maps at the two scales are so distinct suggests that both images should be seen simultaneously, as indeed they are, in transparency.

In other words, we are suggesting that the spurious components affect visibility because they interact with

the signal *within* a given spatial scale, affecting the peaks of local energy. Each scale generates an independent, size-labelled feature map. Spatial correspondence between these symbolic representations at different scales of appropriate contrast produces the perception of an opaque and solid edge that segments the image. For the phase-shifted image, the features at the low-scale do not coincide with that of the block structure, but with the face or letter [Fig. 7(F)]. The features of the high-scale map do not have corresponding features at low-scales, and are therefore perceived as highpass features, in transparency. The low and high spatial frequency maps provide an independent representation, both of which emerge as independent perceptions. This line of reasoning is very similar to that proposed by Marr & Hildreth (1980), suggesting that we do not have perceptual access to the feature maps at each scale, but only after they have been combined. Only if there is disagreement between two scales will the two percepts be seen simultaneously. The major differences between our approaches are the means by which one scale may influence another (see Morrone & Burr, 1993), and in the intrinsic non-linearities in feature marking.

The model simulations explain why the phase shift restores visibility, and manages to predict accurately the conditions under which it does. However, we cannot exclude the possibility that other mechanisms are at work, especially for very blurred images (well below the range considered here), where the local energy model marks very few features. For example, the appearance of simple combinations of low-frequency patterns of different orientation or spatial frequency are not predictable by local energy (Georgeson, 1992, 1994). It is important to emphasize that our model does not assume that the local energy feature map is the only description of the image, but serves as one source of information for *structure and segmentation*. Another important source of information, particularly at very low-scales, could be derived directly from the physical luminance distribution, with very little feature processing (Burr & Morrone, 1994, p. 166; Pessoa *et al.*, 1995). The results of this study suggest that information derived directly from the luminance distribution can be used for recognition purposes, but only in the absence of other competing feature-based information that will dictate image segmentation.

We have chosen to use the local energy model of feature detection to model the present results, as it has been implemented in two dimensions, and readily lends itself to quantitative analysis. However, it is quite feasible that other feature-based models using the principles of MIRAGE (Watt & Morgan, 1985) or MIDAAS (Kingdom & Moulden, 1992) would also work. For example, the rules for synthesizing the symbolic description of a solid edge from separate descriptions at different scales are very similar to those that form the basis of MIDAAS (and quite different from that of Marr & Hildreth, 1980). However, the good qualitative and quantitative performance of our own model is most encouraging.

The nature of transparency

One of the more interesting facts to emerge from this work is that changing image phase often caused the sensation of transparency. Transparency has been studied extensively by Metelli (1970, 1974) earlier this century, and more recently by Adelson (1993) and others. However, almost all studies have employed simple smooth unpatterned stimuli, such as celluloid paper, or computer equivalents, from which a clear set of rules has been developed to predict when surfaces will appear to be transparent or otherwise. The transparency discussed here is somewhat more complicated, where two patterns are seen simultaneously in the same position, one of them appearing transparent.

We would like to suggest a simple explanation for the transparency phenomena reported in this paper. Transparency can be explained by the synthesis of the symbolic feature maps. When an edge marked at high-scales corresponds in location and amplitude with one marked at low-scales, it is labelled as opaque. Only one object is perceived at that location. Otherwise, if the spatial correspondence is broken, so that only the edge at high-scales exists, it will be perceived as a transparent edge. The regions are still seen as constant brightness (like the Craik–O'Brien edge), but any other patterning in the region will be perceived simultaneously. This possibility is currently being investigated with more traditional images, such as those used by Metelli (1974).

REFERENCES

Adelson, E. H. (1993). Perceptual organization and the judgement of brightness. *Science, 262*, 2042–2044.

Burbeck, C. A. (1986). Negative arfterimages and photopic luminance adaptation in human vision. *Journal of the Optical Society of America, A3*, 1159–1165.

Burr, D. C. (1981). Temporal summation of moving images by the human visual system. *Proceedings of the Royal Society of London, B211*, 321–339.

Burr, D. C. & Morrone, M. C. (1990). Edge detection in biological and artificial visual systems. In Blakemore, C. (Ed.), *Vision: coding and efficiency* (pp. 185–194). Cambridge, U.K.: Cambridge University Press.

Burr, D. C. & Morrone, M. C. (1992). A non-linear model of feature detection. In Pinter R. B. & Nabet, B. (Eds), *Non-linear vision* (pp. 309–328). Boca Raton, FL: CRC Press.

Burr, D. C. & Morrone, M. C. (1994) The role of features in constructing visual images. In Morgan, M. J. (Ed.), *Higher-order processing in the visual system* (pp. 129–146). London: John Wiley.

Burr, D. C., Morrone, M. C. & Ross, J. (1986). Local and global visual analysis. *Vision Research, 26*, 749–757.

Campbell, F. W. & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *Journal of Physiology (London), 187*, 437–445.

Canny, J. F. (1983). Finding edges and lines in images. *MIT AI Laboratory Technical Report, 720*

Canny, J. F. (1986). A computational approach to edge detection. *IEEE Transactions PAMI, 8*, 679–698.

Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America, A 5*, 1986–2007.

Costen, N., Parker, D. M. & Craw, I. (1994). Spatial content and statial quantisation effects in face recognition. *Perception, 23*, 129–146.

Costen, N. P., Parker, D. M. & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face recognition. *Perception and Psychophysics, 58*, 602–612.

De Valois, K. K., De Valois, R. L. & Yund, E. W. (1979). Responses of striate cortex cells to grating and checkerboard patterns. *Journal of Physiology (London), 291*, 483–505.

Fleet, D. J. & Langley, K. (1994). Computational analysis of non-Fourier motion. *Vision Research, 34*, 3057–3079.

Georgeson, M. A. (1992). Human vision combines oriented filters to compute edges. *Proceedings of the Royal Society of London, B249*, 235–245.

Georgeson, M. (1994). From filters to features: location, orientation, contrast and blur. In Morgan, M. J. (Ed.), *Higher order processing in the visual system* (pp. 147–165). New York: John Wiley.

Harmon, L. D. (1973). The recognition of faces. *Scientific American, 2295*, 70–83.

Harmon, L. D. & Julesz, B. (1973). Masking in visual recognition: effect of two-dimensional filtered noise. *Science, 180*, 1194–1197.

Kingdom, F. & Moulden, B. (1992). A multi-channel approach to brightness coding. *Vision Research, 32*, 1565–1582.

Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics, 50*, 363–370.

Landy, M. S., Cohen, Y. & Sperling, G. (1984). HIPS: Image processing under UNIX. Software and applications. *Behavioural Research Methods, Instruments and Computing, 16*, 199–216.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Marr, D. & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, B207*, 187–217.

Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. In *Contemporary problems in perception*. London: Taylor and Francis.

Metelli, F. (1974). The perception of transparency. *Scientific American, 230*, 91–98.

Morgan, M. J. & Watt, R. J. (1984). Spatial frequency interference effects and interpolation in vernier acuity. *Vision Research, 24*, 1911–1984.

Morrone, M. C. & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London, B235*, 221–245.

Morrone, M. C. & Burr, D. C. (1993). A model of human feature detection based on matched filters. In Dario, P., Sandini G. & Aebischer, P. (Eds), *Robots and biological systems: towards a new bionics?* (pp. 43–64). Berlin: Springer.

Morrone, M. C. & Burr, D. C. (1994). Visual capture and transparency in blocked images. *Perception, 23S*, 20b.

Morrone, M. C. & Burr, D. C. (1995). Visual transparency depends on phase congruence. *Investigative Ophthalmology and Visual Science Supplement, 36*, S439.

Morrone, M. C., Burr, D. C. & Ross, J. (1983). Added noise restores recognition of coarse quantised images. *Nature, 305*, 226–228.

Nelder, J. A. & Mead, R. (1964). A simplex method for function minimization. *Computer Journal, 7*, 308–313.

Perona, P. & Malik, J. (1990). Detecting and localizing edges composed of steps, peaks and roofs. In *Proceedings of the international conference on computer vision, Osaka, Japan, 1990* (pp. 52–57). New York: Institute of Electrical and Electrical Engineers.

Pessoa, L., Mingolla, E. & Neumann, H. (1995). A contrast- and luminance-driven multiscale network model of brightness perception. *Vision Research, 15*, 2201–2224.

Watt, R. J. & Morgan, M. J. (1982). Mechanisms of interpolation in human spatial vision. *Nature, 299*, 553–555.

Watt, R. J. & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25*, 1661–1667.

Witkin, A. P. (1983). Scale-space filtering. *Proceedings of IJCAI*, 1019-1021.

Acknowledgements—We thank Professor Mark Georgeson for helpful comments on the manuscript. The research was partly supported by target grant CNR ROBOTICA 93.00926.PF67.