

# Why is it easier to identify someone close than far away?

GEOFFREY R. LOFTUS

*University of Washington, Seattle, Washington*

and

ERIN M. HARLEY

*University of California, Los Angeles, California*

It is a matter of common sense that a person is easier to recognize when close than when far away. A possible explanation for why this happens begins with two observations. First, the human visual system, like many image-processing devices, can be viewed as a spatial filter that passes higher spatial frequencies, expressed in terms of cycles/degree, progressively more poorly. Second, as a face is moved farther from the observer, the face's image spatial frequency spectrum, expressed in terms of cycles/face, scales downward in a manner inversely proportional to distance. An implication of these two observations is that as a face moves away, progressively lower spatial frequencies, expressed in cycles/face—and therefore, progressively coarser facial details—are lost to the observer at a rate that is likewise inversely proportional to distance. We propose what we call the *distance-as-filtering hypothesis*, which is that these two observations are sufficient to explain the effect of distance on face processing. If the distance-as-filtering hypothesis is correct, one should be able to simulate the effect of seeing a face at some distance,  $D$ , by filtering the face so as to mimic its spatial frequency composition, expressed in terms of cycles/face, at that distance. In four experiments, we measured face perception at varying distances that were simulated either by filtering the face as just described or by shrinking the face so that it subtended the visual angle corresponding to the desired distance. The distance-as-filtering hypothesis was confirmed perfectly in two face perception tasks: assessing the informational content of the face and identifying celebrities. Data from the two tasks could be accounted for by assuming that they were mediated by different low-pass spatial filters within the human visual system that have the same general mathematical description but that differ in scale by a factor of approximately 0.75. We discuss our results in terms of (1) how they can be used to explain the effect of distance on visual processing, (2) what they tell us about face processing, (3) how they are related to “flexible spatial scale usage,” as discussed by Schyns and colleagues, and (4) how they may be used in practical (e.g., legal) settings to demonstrate the loss of face information that occurs when a person is seen at a particular distance.

## PROLOGUE

In the year 1997, October 10 was the day on which the citizens of Alaska received their annual oil dividend of approximately \$1,500 per person. Perhaps it was to celebrate this event that four young hoodlums decided to hit the streets of Fairbanks for a *Clockwork-Orange*-style evening of vicious attacks on random individuals. When the night was over, a teenage boy, identified as L.H., lay dead, and an older man, Franklin Dayton, was seriously injured as a result of the gang's marauding.

Fairbanks, Alaska is not a town that takes violent events in stride: The public responded in outrage, and police reaction was swift. Within days, four suspects were arrested and locked up. Eventually, they were tried for the crimes.

At trial, the prosecutor introduced various kinds of evidence bearing on the defendants' guilt. However, the prosecutor had no smoking gun, literal or metaphorical, and the centerpiece of his case was an eyewitness account provided by one Arlo Olson. Mr. Olson testified that while standing in the doorway of Eagles Hall in downtown Fairbanks, he had watched in horror as a group of men, whom he later identified as the defendants, accosted and savagely beat Mr. Dayton in a parking lot a couple of blocks away.

Of some note is that “a couple of blocks away”—the distance from Olson's vantage point to the parking lot—was determined to be approximately 450 ft. In response to this and related issues, the defense flew in an expert witness from Seattle—a psychologist whose job was to

---

This research was supported by NIMH Grant MH41637 to G.R.L. We thank Keri Carlsen and Jacquie Pickrell for assistance in data collection. We thank Davida Teller for detailed and very useful comments on two earlier drafts. For additional insights and comments, we also thank Larry Maloney, Cathleen Moore, John Palmer, Denis Pelli, and Gillian Rhodes. Correspondence concerning this article should be addressed to G. R. Loftus, Box 351525, University of Washington, Seattle, WA 98195-1525 (e-mail: gloftus@u.washington.edu).

educate the jury about various pitfalls of human perception and memory. Part of the expert's task was to provide information about a witness's ability to perceive and acquire, from a distance of 450 ft, sufficient information about an assailant's appearance to allow him to accurately identify the assailant later on.

The eyewitness, Mr. Olson, testified that he had a strong memory of the defendants. Furthermore, he was quite certain that his strong memory came about because he had had a good view of the defendants as he watched the mayhem unfold in the parking lot. The expert did not dispute Mr. Olson's strong memory and entertained several hypotheses about how Mr. Olson may have acquired it. But contrary to Mr. Olson's claim, none of these hypotheses involved the proposition that his memory had been formed during the actual event. The main reason for the expert's skepticism was that 450 ft is simply too far away for a witness to be able to accurately perceive the features that constitute a person's facial appearance. In his trial testimony, the expert tried to convey this difficulty to the jury by pointing out that seeing someone from 450 ft away is what one is doing when one is sitting high in the center field bleachers of Yankee Stadium, looking across the ballpark at another individual sitting in the stands behind home plate.

On his way back to Seattle, the expert pondered how he might better convey the effect of distance on visual information acquisition. He felt that the Yankee Stadium example may have sufficed for the somewhat extreme circumstances of this particular case. But in other instances, a witness may have viewed a criminal from a distance of 200 or 100 or 50 ft. How, the expert wondered, could a jury or anyone else be provided a clear intuition for the information loss that results from seeing someone at successively greater distances, beyond the bland and relatively uninformative assertion that "because of acuity limitations, one's ability to perceive something gets worse as distance increases"? Perhaps, he thought, some kind of visual representation of such information loss could be devised.

And thus were sown the seeds of the research to be described in this article. The general idea we propose is that the effect of distance on face perception can be construed as the visual system's inability to perceive and encode progressively coarser grained facial details as the face moves farther away. The research reported in the remainder of this article had two purposes. The first was to develop and test a formal theory incorporating this idea and to determine empirically whether it makes sense. The second was to begin to develop quantitative tools for demonstrating the loss of facial detail that corresponds to any specified distance.

### DISTANCE, VISUAL ANGLE, AND SPATIAL FREQUENCIES

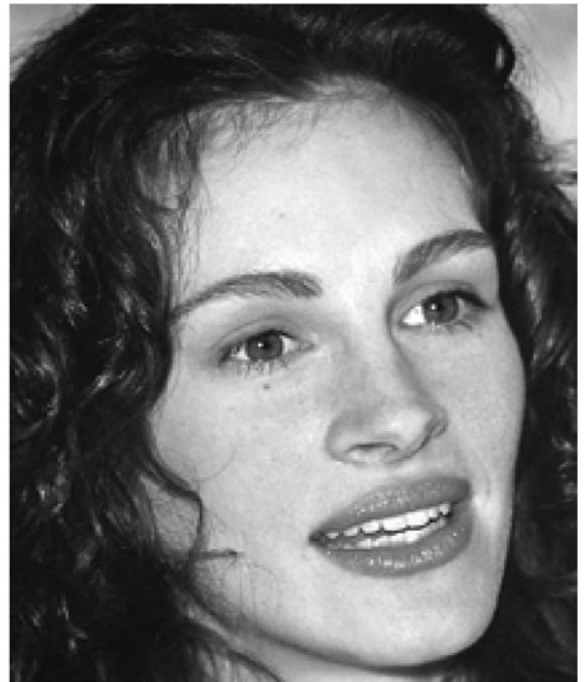
The proposition that "the visual system is less able to perceive and encode progressively coarser grained facial

details as the face moves further away" is vague. It needs to be expressed more precisely before it can be used to explain the effect of distance on perception. In this section, such precision is developed. We will begin by noting the relation between distance and visual angle and will progress from there to the relation between distance and spatial frequency. The section ends with a specific, quantitative hypothesis, called the *distance-as-filtering hypothesis*, that is suitable for experimental test.

### Distance and Visual Angles

If you ask why a face is harder to recognize when it is farther away, most people will answer, "because it gets smaller." This is obviously true in the sense that as a face is moved farther from an observer, it shrinks from the observer's perspective; that is, its visual angle and retinal size decrease in a manner that is inversely proportional to distance. So if one wanted to demonstrate the effect of distance on perception, one could do so by appropriately shrinking a picture. This is done in Figure 1, which shows

5.4 ft



43 ft



172 ft



**Figure 1.** Representation of distance by reduction of visual angle. The visual angles implied by the three viewing distances are correct if this page is viewed from a distance of 22 in.

the size that Julia Roberts's face would appear to be, as seen from distances of 5.4, 43, and 172 ft.

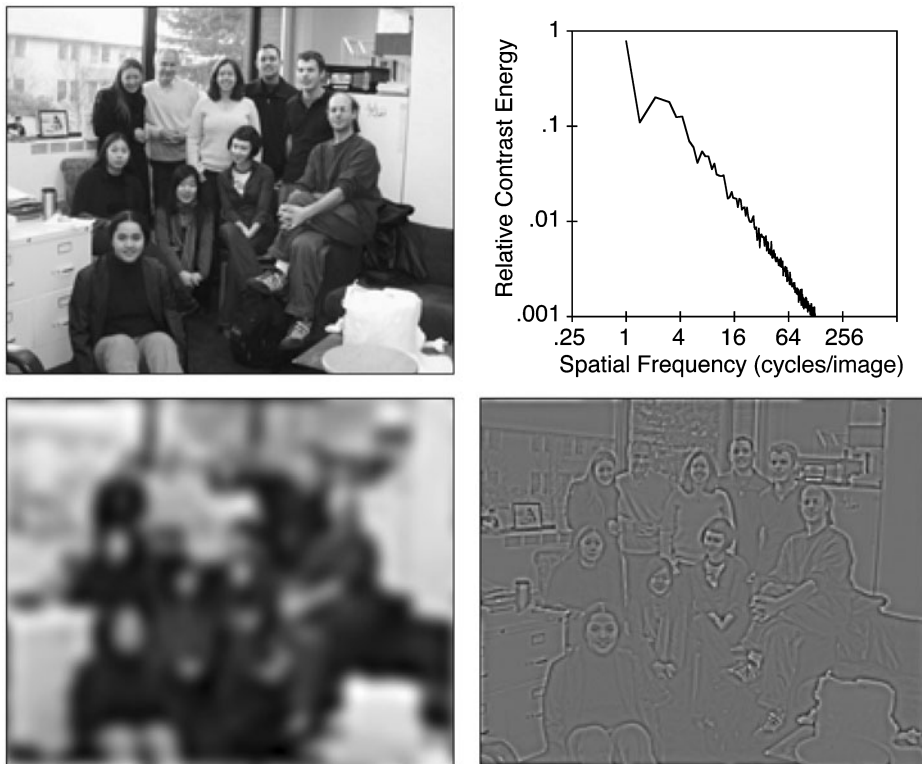
The demonstration in Figure 1 of the effect of distance on perception is unsatisfying for several reasons. From a practical perspective, it is unsatisfying because (1) its quantitative validity depends on the viewer's being a specific distance from the display medium (in this instance, 22 in. from the paper on which Figure 1 is printed) and (2) the graininess of the display medium can differentially degrade the different-sized images—for instance, if one is viewing an image on a computer monitor, an image reduction from, say,  $500 \times 400$  pixels, representing a 5-ft distance, to  $10 \times 8$  pixels, representing a 250-ft distance, creates an additional loss of information, due to pixel reduction, above and beyond that resulting from simply scaling down the image's size (and this is true with any real-life display medium: Looking at the bottom image of Figure 1 through a magnifying glass, for instance, will never improve its quality to the point where it looks like the top image). From a scientific perspective, the demonstration in Figure 1 is unsatisfying because reducing an image's size is simply geometry. That is, apart from the basic notion of a retinal image, it does not use any known information about the visual

system, nor does it provide any new insight about how the visual system works.

### Spatial Frequencies

An alternative to using variation in an image's size to represent distance is to use variation in the image's *spatial frequency* composition. It is well known that any image can be represented equivalently in *image space* and *frequency space*. The image space representation is the familiar one: It is simply a matrix of pixels differing in color or, as in the example shown in Figure 1, gray scale values. The less familiar frequency space representation is based on the well-known theorem that any two-dimensional function, such as the values of a matrix, can be represented as a weighted sum of sine wave gratings of different spatial frequencies and orientations at different phases (Bracewell, 1986). One can move back and forth between image space and frequency space via Fourier transformations and inverse Fourier transformations.

Different spatial frequencies carry different kinds of information about an image. To illustrate, the bottom panels of Figure 2 show the top-left image dichotomously partitioned into its low spatial frequency components (bottom left) and high spatial frequency components (bottom right).



**Figure 2.** Decomposition of a naturalistic scene (top left) into low spatial frequency components (bottom left) and high spatial frequency components (bottom right). The top right panel shows the contrast energy spectrum of the top-left picture.

nents (bottom right). It is evident that the bottom-left picture carries a global representation of the scene, whereas the bottom-right picture conveys information about edges and details, such as the facial expressions and the objects on the desk. Figure 2 provides an example of the general principle that higher spatial frequencies are better equipped to carry information about fine details. In general, progressively lower spatial frequencies carry information about progressively coarser features of the image.

Of interest in the present work is the image's *contrast energy spectrum*, which is the function relating contrast energy—informally, the amount of *presence* of some spatial frequency component in the image—to spatial frequency. For natural images, this function declines and is often modeled as a power function,  $E = kf^{-r}$ , where  $E$  is contrast energy,  $f$  is spatial frequency, and  $k$  and  $r$  are positive constants; thus, there is less energy in the higher image frequencies. Figure 2, upper right panel, shows this function, averaged over all orientations, for the upper left image (note that, characteristically, it is roughly linear on a log-log scale). Below, we will describe contrast energy spectra in more detail, in conjunction with our proposed hypothesis about the relation between distance and face processing.

### Absolute Spatial Frequencies and Image Spatial Frequencies

At this point, we will explicate a distinction between two kinds of spatial frequencies that will be critical in our subsequent logic. *Absolute spatial frequency* is defined as spatial frequency measured in cycles per degree of visual angle (cycles/deg). *Image spatial frequency* is defined as spatial frequency measured in cycles per image (cycles/image). Notationally, we will use  $F$  to denote absolute spatial frequency ( $F$  in cycles/deg) and  $f$  to denote image spatial frequency ( $f$  in cycles/image).

Note that the ratio of these two spatial frequency measures,  $F/f$  in image/degree, is proportional to observer–stimulus distance. Imagine, for instance, a stimulus consisting of a piece of paper, 1 m high, depicting 10 cycles of a horizontally aligned sine wave grating. The image frequency would thus be  $f = 10$  cycles/image. If this stimulus were placed at a distance of 57.29 m from an observer, its vertical visual angle could be calculated to be  $1^\circ$ , and the absolute spatial frequency of the grating would, therefore, be  $F = 10$  cycles/1 deg = 10 cycles/deg—that is,  $F/f = (10 \text{ cycles/deg})/(10 \text{ cycles/image}) = 1$  image/deg. If the observer–stimulus distance were increased—say, by a factor of 5—the visual angle would be reduced to  $1/5 = 0.2^\circ$ , the absolute spatial frequency would be increased to  $F = 10 \text{ cycles}/0.2 \text{ deg} = 50 \text{ cycles/deg}$ , and  $F/f = 50/10 = 5$  image/deg. If, on the other hand, the distance were *decreased*—say, by a factor of 8—the visual angle would be increased to  $1 \times 8 = 8^\circ$ , the absolute spatial frequency would be  $F = 10 \text{ cycles}/8 \text{ deg} = 1.25 \text{ cycles/deg}$ , and  $F/f = 1/8$  image/deg. And so on.

### Filters, Modulation Transfer Functions, and Contrast Sensitivity Functions

Central to our ideas is the concept of a *spatial filter*, which is a visual processing device that differentially passes different spatial frequencies. A filter's behavior is characterized by a *modulation transfer function* (MTF), which assigns an *amplitude scale factor* to each spatial frequency. The amplitude scale factor ranges from 1.0 for spatial frequencies that are completely passed by the filter to 0.0 for spatial frequencies that are completely blocked by it.

The human visual system can be construed as consisting of a collection of components—for example, the optics of the eye, the receptive fields of retinal ganglion cells, and so on—each component acting as a spatial filter. This collection of filters results in an overall MTF in humans, whose measured form depends on the particular physical situation and the particular task in which the human is engaged.

What do we know about the human MTF? In certain situations, the MTF can be described by a *contrast sensitivity function* (CSF), which is the reciprocal of threshold contrast (i.e., contrast sensitivity) as a function of absolute spatial frequency. A “generic” CSF, shown in Figure 3, reasonably resembles those obtained empirically under static situations (e.g., Campbell & Robson, 1968; van Nes & Bouman, 1967). As can be seen, it is *band-pass*; that is, it best represents spatial frequencies around 3 cycles/deg, whereas both lower and higher spatial frequencies are represented more poorly. However, under a variety of conditions, the CSF is low-pass rather than band-pass. These conditions include (1) low luminance (e.g., van Nes & Bouman, 1967), (2) high contrast (Georgeson & Sullivan, 1975), and (3) temporally varying, rather than static, stimuli (e.g., Robson, 1966).

Thus, the form of the CSF under a variety of circumstances is known. However, measurements of the CSF have been carried out using very simple stimuli—typically, sine wave gratings—under threshold conditions,

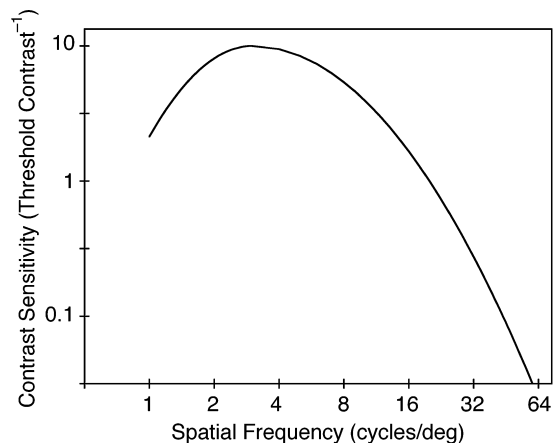


Figure 3. Band-pass approximation to a human contrast sensitivity function for a low-contrast, high-luminance, static scene.

and its application to suprathreshold complex stimuli, such as faces, is dubious. One of the goals of the present research was to estimate at least the general form of the relevant MTF for the face-processing tasks with which we are concerned.

**Acuity.** Whereas measurement of the CSF has been a topic of intense scrutiny among vision scientists, practitioners in spatial vision (e.g., optometrists) rely mainly on measurement of *acuity*. As is known by anyone who has ever visited an eye doctor, acuity measurement entails determining the smallest stimulus of some sort—“smallest” defined in terms of the visual angle subtended by the stimulus—that one can correctly identify. Although the test stimuli are typically letters, acuity measurements have been carried out over the past century with numerous stimulus classes, including line separation (e.g., Shlaer, 1937) and Landolt C gap size (e.g., Shlaer, 1937). As has frequently been pointed out (e.g., Olzak & Thomas, 1989, pp. 7–45), measurement of acuity essentially entails measurement of a single point on the human MTF—namely, the high-frequency cutoff—under suprathreshold (high-luminance and high-contrast) conditions.

The issues that we raised in our Prologue and the research that we will describe are intimately concerned with the question of how close a person must be—that is, how large a visual angle the person must subtend—to become recognizable. In other words, we are concerned with acuity. However, we are interested in more than acuity, for two reasons. First, our goals extend beyond simply identifying the average distance at which a person can be recognized. We would like, instead, to be able to specify what, from the visual system’s perspective, is the spatial frequency composition of a face at any given distance. Achievement of such goals would constitute both scientific knowledge necessary for building a theory of how visual processing depends on distance and practical knowledge useful for conveying an intuitive feel for distance effects to interested parties, such as juries. Second, it is known that face processing in particular depends on low and medium spatial frequencies, not just on high spatial frequencies (e.g., Harmon & Julesz, 1973).

### Distance and Spatial Frequencies

Returning to the issue of face perception at a distance, suppose that there were no drop in the MTF at lower spatial frequencies. (Below, I will argue that, for purposes of the present applications, this supposition is plausible.) This would make the corresponding MTF *low-pass*, rather than *band-pass*.

Let us, for the sake of illustration, assume a low-pass MTF that approaches zero around  $F = 30$  cycles/deg (a value estimated from two of the experiments described below); that is, absolute spatial frequencies above 30 cycles/deg can be thought of as essentially invisible to the visual system. Focusing on this 30 cycles/deg upper limit, we consider the following example. Suppose that a face is viewed from 43 ft away, at which distance it subtends a visual angle of approximately  $1^\circ$ .<sup>1</sup> This means

that at this particular distance, image frequency  $f$  equals absolute frequency  $F$  and therefore spatial frequencies greater than about  $f = 30$  cycles per *face* will be invisible to the visual system; that is, facial details smaller than about  $1/30$  of the face’s extent will be lost. Now suppose the distance is increased—say, by a factor of 4—to 172 ft. At this distance, the face will subtend approximately  $0.25^\circ$  of visual angle, and the spatial frequency limit of  $F = 30$  cycles/deg translates into  $f = 30 \text{ cycles/deg} \times 1/4 \text{ deg/face} = 7.5 \text{ cycles/face}$ . Thus, at a distance of 172 ft, details smaller than about  $(1/30) \times 4 = 2/15$  of the face’s extent will be lost. If the distance is increased by a factor of 10—to 430 ft—details smaller than about  $(1/30) \times 10 = 1/3$  of the face’s extent will be lost. And so on. In general, progressively coarser facial details are lost to the visual system in a manner that is inversely proportional to distance.

These informal observations can be developed into a more complete mathematical form. Let the filter corresponding to the visual system’s MTF be expressed as

$$A = c(F), \quad (1)$$

where  $A$  is the amplitude scale factor,  $F$  is frequency in cycles/degree, and  $c$  is the function that relates the two. Now consider  $A$  as a function not of  $F$ , absolute spatial frequency, but of image spatial frequency,  $f$ . Because 43 ft is the approximate distance at which a face subtends  $1^\circ$ , it is clear that  $f = F * (43/D)$ , where  $D$  is the face’s distance from the observer. Therefore, the MTF defined in terms of cycles/face is

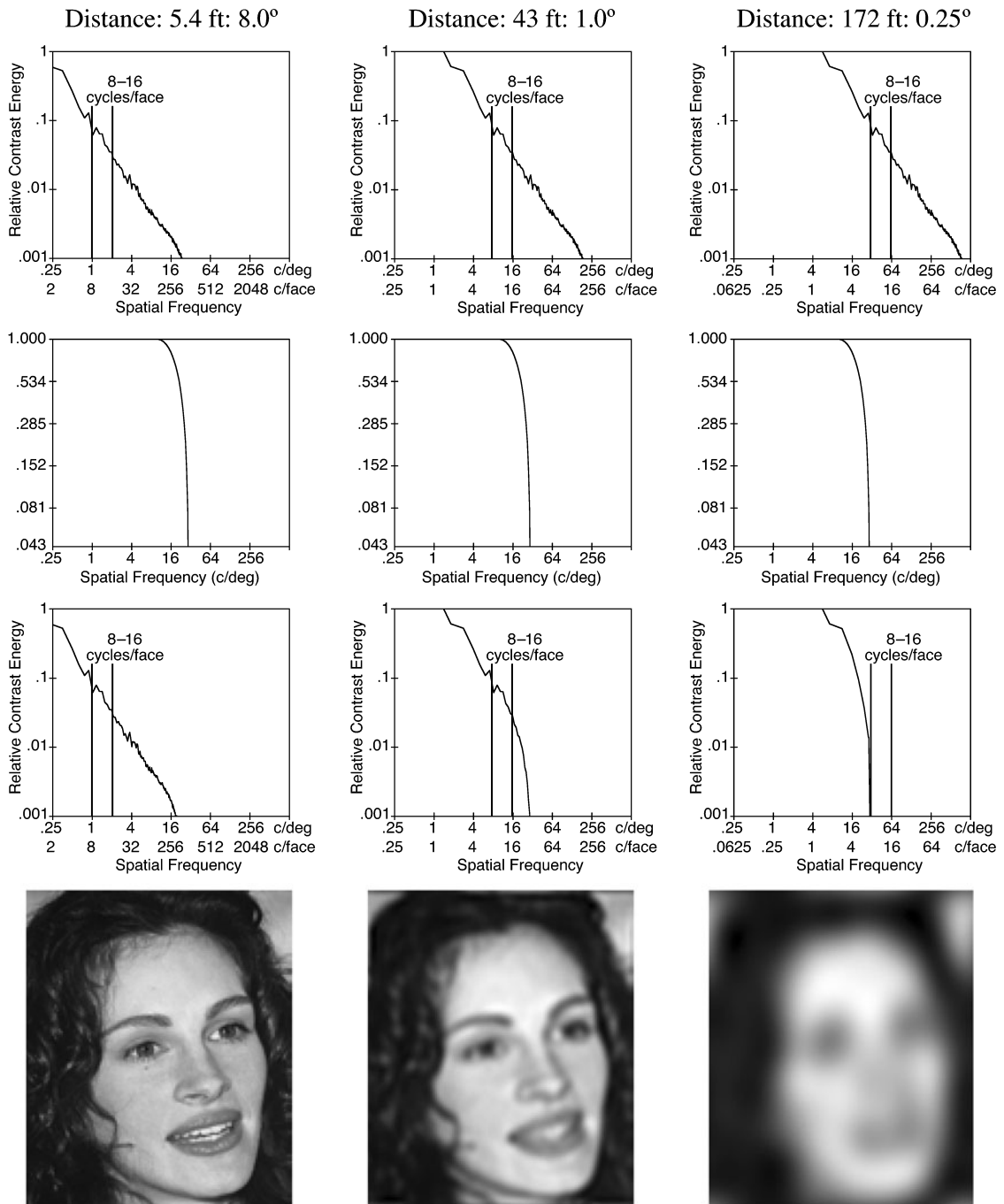
$$A = c(f) = c\left(\frac{43F}{D}\right). \quad (2)$$

Simple though it is, Equation 2 provides, as we shall see, the mathematical centerpiece of much of what follows.

### Distance Represented by Filtering

This logic implies an alternative to shrinking a face (Figure 1) as a means of representing the effect of distance. Assume that one knows the visual system’s MTF for a particular set of circumstances—that is, that one knows the  $c$  in Equations 1 and 2. Then, to represent the face as seen by an observer at any given distance  $D$ , one could remove high image frequencies by constructing a filter as defined by Equation 2 and applying the filter to the face.

Figure 4 demonstrates how this is done. Consider first the center panels, in which Julia Roberts is assumed to be viewed from a distance of  $D = 43$  ft (at which point, recall,  $F = f$ ). We began by computing the Fourier transform of the original picture of her face. The resulting contrast energy spectrum, averaged over all orientations, is shown in the top center panel (again, roughly linear on a log-log scale). Note that the top-row (and third-row) panels of Figure 4 have two abscissa scales: absolute frequency in cycles/degree and image frequency in cycles/face. To provide a reference point, we have indicated the region that contains between 8 and 16 cycles/face, an approximation of which has been suggested as being impor-



**Figure 4.** Demonstration of low-pass filtering and its relation to distance. The columns represent three distances ranging from 5.4 to 172 ft. Top row: contrast energy spectrum at the three distances (averaged across orientations). Second row: Assumed low-pass modulation transfer function corresponding to the human visual system. Third row: Result of multiplying the filter by the spectrum: With longer distances, progressively lower image frequencies are eliminated. Bottom row: Filtered images—the phenomenological appearances at the various distances—that result. c/deg, cycles per degree; c/face, cycles per face.

tant for face recognition (e.g., Morrison & Schyns, 2001). Because at  $D = 43$  ft the face subtends, as has been indicated earlier,  $1^\circ$  of visual angle, 8–16 cycles/face corresponds to 8–16 cycles/deg. The second row shows a low-

pass filter: a candidate human MTF. We will describe this filter in more detail below, but essentially, it passes absolute spatial frequencies perfectly up to 10 cycles/deg and then falls parabolically, reaching zero at 30 cycles/

deg. The third row shows the result of filtering, which, in frequency space, entails simply multiplying the top-row contrast energy spectrum by the second-row MTF on a point-by-point basis. Finally, the bottom row shows the result of inverse-Fourier transforming the filtered spectrum (i.e., the third-row spectrum) back to image space. The result—the bottom middle image—is slightly blurred, in comparison with the original (Figure 1), because of the loss of high spatial frequencies expressed in terms of cycles/face (compare the top-row original spectrum to the third-row filtered spectrum).

Now consider the left panels of Figure 4. Here, Ms. Roberts is presumed to be seen from a distance of 5.4 ft; that is, she has moved closer by a factor of eight and, thus, subtends a visual angle of  $8^\circ$ . This means that the Fourier spectrum of her face has likewise scaled down by a factor of eight; thus, for instance, the 8–16 cycles/face region now corresponds to 1–2 cycles/deg. At this distance, the MTF has had virtually no effect on the frequency spectrum; that is, the top-row unfiltered spectrum and the third-row filtered spectrum are virtually identical and as a result, the image is unaffected. Finally, in the right panel, Ms. Roberts has retreated to 172 ft away, where she subtends a visual angle of  $0.25^\circ$ . Now her frequency spectrum has shifted *up*, so that the 8–16 cycles/face corresponds to 32–64 cycles/deg. Because the MTF begins to descend at 10 cycles/deg and obliterates spatial frequencies greater than 30 cycles/deg, much of the high spatial frequency information has vanished. The 8–16 cycles/face information, in particular, has been removed. As a result, the filtered image is very blurred—arguably, to the point that one can no longer recognize whom it depicts.

### Rationale for a Low-Pass MTF

To carry out the kind of procedure just described, it was necessary to choose a particular MTF—that is, to specify the function  $c$  in Equations 1 and 2. Earlier, we asserted that, for present purposes, it is reasonable to ignore the MTF falloff at lower spatial frequencies. As we will describe in more detail below, the assumed MTFs used in the calculations to follow are low-pass. There are three reasons for this choice.

**Suprathreshold contrast measurements.** Although there are numerous measurements of the visual system's threshold CSF, there has been relatively little research whose aim has been to measure the suprathreshold CSF. One study that did report such work was that by Georgeson and Sullivan (1975). They used a matching paradigm in which observers adjusted the perceived contrast of a comparison sine wave grating shown at one of a number of spatial frequencies to the perceived contrast of a 5 cycles/deg test grating. They found that at low contrasts, the resulting functions relating perceived contrast to spatial frequency were band-pass. At higher contrast levels (beginning at a contrast of approximately 0.2), the functions were approximately flat between 0.25 and 25 cycles/deg, which were the limits of the spatial frequency values that the authors reported.

**Absolute versus image frequency: Band-passed faces at different viewing distances.** Several experiments have been reported in which absolute and image spatial frequency have been disentangled using a design in which image spatial frequency and observer–stimulus distance—which influences absolute spatial frequency—have been factorially varied (Hayes, Morrone, & Burr, 1986; Parish & Sperling, 1991). Generally, these experiments indicate that image frequency is important and robust in determining various kinds of task performance, whereas absolute frequency makes a difference only with high image frequencies.

To illustrate the implication of such a result for the shape of the MTF, consider the work of Hayes et al. (1986) who reported a face identification task: Target faces, band-passed at various image frequencies, were presented to observers, who attempted to identify them. The band-pass filters were 1.5 octaves wide and were centered at one of five image frequency values ranging from 4.4 to 67 cycles/face. The faces appeared at one of two viewing distances, differing by a factor of four—which, of course, meant that the corresponding absolute spatial frequencies also differed by a factor of four. Hayes et al. found that recognition performance depended strongly on image frequency but depended on viewing distance only with the highest image frequency filter—that is, the one centered at 67 cycles/face. This filter passed absolute spatial frequencies from approximately 1.7 to 5.0 cycles/deg for the short viewing distance and from approximately 7 to 20 cycles/deg for the long viewing distance. Performance was better at the short viewing distance, corresponding to the lower absolute spatial frequency range, than at the longer distance, corresponding to the higher absolute spatial frequency range. The next highest filter, which produced very little viewing distance effect, passed absolute spatial frequencies from approximately 0.9 to 2.5 cycles/deg for the short viewing distance and from approximately 3.5 to 10 cycles/deg for the long viewing distance.

These data can be explained by the assumptions that (1) performance is determined by the visual system's representation of spatial frequency in terms of cycles/face and (2) the human MTF in this situation is low-pass—it passes absolute spatial frequencies perfectly up to some spatial frequency between 10 and 20 cycles/deg, before beginning to drop. Thus, with the second highest filter, the perceptible frequency range would not be affected by the human MTF at either viewing distance. With the highest filter, however, the perceptible frequency range would be affected by the MTF at the long, but not at the short, viewing distance.

**Phenomenological considerations.** Suppose that the suprathreshold MTF were band-pass as in Figure 3. In that case, we could simulate what a face would look like at varying distances, just as we described doing earlier using a low-pass filter. In Figure 5, which is organized like Figure 4, we have filtered Julia Roberts's face with a band-pass filter, centered at 3 cycles/deg. It is constructed such that it falls to zero at 30 cycles/deg, just as





ical results for long simulated distances, but at short simulated distances, yields images that look very different than actual objects seen close up. A low-pass filter, in contrast, yields images that appear phenomenologically reasonable at all virtual distances.

### The Distance-As-Filtering Hypothesis

We will refer to the general idea that we have been describing as the *distance-as-filtering hypothesis*. Specifically, the distance-as-filtering hypothesis is the conjunction of the following two assumptions: (1) The difficulty in perceiving faces at increasing distances comes about because the visual system's limitations in representing progressively lower image frequencies, expressed in terms of cycles/face, cause a loss of increasingly coarser facial details, and (2) if an appropriate MTF  $c$  (see Equations 1 and 2 above) can be determined, the representation of a face viewed from a particular distance,  $D$ , is equivalent to the representation acquired from the version of the face that is filtered in accord with Equation 2.

### The Notion of Equivalence

Before proceeding, we would like to clarify what we mean by *equivalent*. There are examples within psychology wherein physically different stimuli give rise to representations that are genuinely equivalent at any arbitrary level of the sensory-perceptual-cognitive system, because the information that distinguishes the stimuli is lost at the first stage of the system. A prototypical example is that of color metamers—stimuli that, although physically different wavelength mixtures, engender identical quantum-catch distributions in the photoreceptors. Because color metamers are equivalent at the photoreceptor stage, they must therefore be equivalent at any subsequent stage.

When we characterize distant and low-pass filtered faces as *equivalent*, we do not, of course, mean that they are equivalent in this strong sense. They are not phenomenologically equivalent: One looks small and clear, the other looks large and blurred, and they are obviously distinguishable. One could, however, propose a weaker definition of *equivalence*. In past work, we have suggested the term *informational metamers* in reference to two stimuli that, although physically and phenomenologically different, lead to presumed equivalent representations *with respect to* some task at hand (see Harley, Dillon, & Loftus, 2004; Loftus & Ruthruff, 1994). For instance Loftus, Johnson, and Shimamura (1985) found that a  $d$ -msec unmasked stimulus (that is a stimulus plus an icon) is equivalent to a  $(d + 100)$ -msec masked stimulus (a stimulus without an icon) with respect to subsequent memory performance across a wide range of circumstances. Thus, it can be argued that the representations of these two kinds of physically and phenomenologically different stimuli eventually converge into equivalent representations at some point prior to whatever representation underlies task performance.

Similarly, by the distance-as-filtering hypothesis, we propose that reducing the visual angle of the face, on the one hand, and appropriately filtering the face, on the other hand, lead to representations that are equivalent in robust ways with respect to performance on various tasks. In the experiments that we will report below, we confirmed such equivalence with two tasks.

### Different Absolute Spatial Frequency Channels?

A key implication of the distance-as-filtering hypothesis is that the visual system's representation of a face's image frequency spectrum is necessary and sufficient to account for distance effects on face perception: That is, two situations—a distant (i.e., small retinal image) unfiltered face and a closer (i.e., large retinal image) suitably filtered face—will produce functionally equivalent representations and, therefore, equal performance.

Note, however, that for these two presumably equivalent stimuli, the same image frequencies correspond to different absolute frequencies: They are higher for the small unfiltered face than for the larger filtered face. For instance, a test face sized to simulate a distance of 108 ft would subtend a visual angle of approximately  $0.40^\circ$ . A particular image spatial frequency—say, 8 cycles/face—would therefore correspond to an absolute spatial frequency of approximately 20 cycles/deg. A corresponding large filtered face, however, subtends, in our experiments, a visual angle of approximately  $20^\circ$ , so the same 8 cycles/face would correspond to approximately 0.4 cycles/deg.

There is evidence from several different paradigms that the visual system decomposes visual scenes into separate spatial frequency channels (Blakemore & Campbell, 1969; Campbell & Robson, 1968; De Valois & De Valois, 1980, 1988; Graham, 1989; Olzak & Thomas, 1986). If this proposition is correct, it would mean that two presumably equivalent stimulus representations—a small unfiltered stimulus, on the one hand, and a large filtered stimulus, on the other—would issue from different spatial frequency channels. One might expect that the representations would thereby not be equivalent in any sense—that is, that the distance-as-filtering hypothesis would fail under experimental scrutiny. As we shall see, however, contrary to such an expectation, the hypothesis holds up quite well.

### General Prediction

With this foundation, a general prediction of the distance-as-filtering hypothesis can be formulated: It is that in any task requiring visual face processing, performance for a face whose distance is simulated by appropriately sizing it will equal performance for a face whose distance is simulated by appropriately filtering it.

## EXPERIMENTS

We will report four experiments designed to test this general prediction. In Experiment 1, observers matched the informational content of a low-pass-filtered *com-*

parison stimulus to that of a variable-sized test stimulus. In Experiments 2–4, observers attempted to recognize pictures of celebrities that were degraded by either low-pass filtering or size reduction.

### Experiment 1 Matching Blur to Size

Experiment 1 was designed to accomplish two goals. The first was to provide a basic test of the distance-as-filtering hypothesis. The second goal, given reasonable accomplishment of the first, was to begin to determine the appropriate MTF for representing distance by spatial filtering. In quest of these goals, a *matching paradigm* was devised. The observers viewed an image of a test face presented at one of six different sizes. Each size corresponded geometrically to a particular observer–face distance,  $D$ , that ranged from 20 to 300 ft. For each test size, the observers selected which of 41 progressively more blurred comparison faces best matched the perceived informational content of the test face.

The set of 41 comparison faces was constructed as follows. Each comparison face was generated by low-pass filtering the original face, using a version of Equation 2 to be described in detail below. Across the 41 comparison faces, the filters removed successively more high spatial frequencies, and the faces became, accordingly, more and more blurred. From the observer's perspective, these faces ranged in appearance from completely clear (when the filter's spatial frequency cutoff was high and it thereby removed relatively few high spatial frequencies) to extremely blurry (when the filter's spatial frequency cutoff was low and it thereby removed most of the high spatial frequencies). For each test face size, the observer, who was permitted complete untimed access to all 41 comparison faces, selected the particular comparison face that he or she felt best matched the perceived informational content of the test face. Thus, in general, a large (simulating a close) test face was matched by a relatively clear comparison face, whereas a smaller (simulating a more distant) test face was matched by a blurrier comparison face.

### Constructing the Comparison Faces

In this section, we will provide the quantitative details of how the 41 filters were constructed in order to generate the corresponding 41 comparison faces.

We have already argued that a low-pass filter is appropriate as a representation of the human MTF in this situation and, thus, as a basis for the comparison pictures to be used in this task. A low-pass filter can take many forms. Somewhat arbitrarily, we chose a filter that is constant at 1.0 (which means that it passes spatial frequencies perfectly) up to some *rolloff* spatial frequency, termed  $F_0$  cycles/deg, then declines as a parabolic function of log spatial frequency, reaching zero at some *cutoff* spatial frequency, termed  $F_1$  cycles/deg, and remaining at zero for all spatial frequencies greater than  $F_1$ . We in-

roduced a constant,  $r > 1$ , such that  $F_0 = F_1/r$ . Note that  $r$  can be construed as the relative slope of the filter function: Lower  $r$  values correspond to steeper slopes.

Given this description, for *absolute* spatial frequencies defined in terms of  $F$  cycles/deg, the filter is completely specified and is derived to be

$$c(F) = \begin{cases} 1.0 & \text{for } F < F_0 \\ 1 - \left[ \frac{\log(F/F_0)}{\log(r)} \right]^2 & \text{for } F_0 \leq F \leq F_1 \\ 0.0 & \text{for } F > F_1. \end{cases} \quad (3)$$

Above, we noted that *image* spatial frequency expressed in terms of  $f$ , frequency in cycles/face, is  $f = (43/D)*F$ , where  $D$  is the observer's distance from the face. Letting  $k = (43/D)$ , the Equation 3 filter expressed in terms of  $f$  is

$$c(f, D) = \begin{cases} 1.0 & \text{for } f < kF_0 \\ 1 - \left[ \frac{\log(f/kF_0)}{\log(r)} \right]^2 & \text{for } kF_0 \leq f \leq kf_1 \\ 0.0 & \text{for } f > kf_1. \end{cases} \quad (4)$$

In Equation 4,  $kF_0$  and  $kF_1$  correspond to what we term  $f_0$  and  $f_1$ , which are, respectively, the rolloff and the cutoff frequencies, defined in terms of cycles/face.

Because this was an exploratory venture, we did not know what value of  $r$  would be most appropriate. For that reason, we chose two somewhat arbitrary values of  $r$ : 3 and 10. The 41 comparison filters constructed for each value of  $r$  were selected to produce corresponding images having, from the observer's perspective, a large range from very clear to very blurry. Expressed in terms of cutoff frequency in cycles/face ( $f_1$ ), the ranges were from 550 to 4.3 cycles/face for the  $r = 10$  comparison filters and from 550 to 5.2 cycles/face for the  $r = 3$  comparison filters. Each comparison image produced by a comparison filter was the same size as the original image ( $1,100 \times 900$  pixels).

To summarize, each comparison face was defined by a value of  $f_1$ . On each experimental trial, we recorded the comparison face—that is, that value of  $f_1$  that was selected by the observer as matching the test face displayed on that trial. Thus, the distance corresponding to the size of the test face was the independent variable in the experiment, and the selected value of  $f_1$  was the dependent variable.

### Prediction

Given our filter construction process, a candidate MTF is completely specified by values of  $r$  and  $F_1$ . As was indicated, we selected two values of  $r$ , 3 and 10. We allowed the cutoff frequency,  $F_1$ , to be a free parameter estimated in a manner to be described below.

Suppose that the distance-as-filtering hypothesis is correct—that seeing a face from a distance is indeed equivalent to seeing a face whose high image frequen-

cies have been appropriately filtered out. In that case, a test face sized to subtend a visual angle corresponding to some distance,  $D$ , should be matched by a comparison face filtered to represent the same distance. As specified by Equation 2 above,  $f_1 = (43F_1)/D$ , or

$$D = \frac{43F_1}{f_1}, \quad (5)$$

which means that

$$\frac{1}{f_1} = \left( \frac{1}{43F_1} \right) D. \quad (6)$$

Equation 6 represents our empirical prediction: The measured value of  $1/f_1$  is predicted to be proportional to the manipulated value of  $D$  with a constant of proportionality equal to  $1/(43F_1)$ . Given that Equation 6 is confirmed,  $F_1$ —and thus, in conjunction with  $r$ , the MTF as a whole—can be estimated by measuring the slope of the function relating  $1/f_1$  to  $D$ , equating the slope to  $1/(43F_1)$ , solving for  $F_1$ , and plugging the resulting  $F_1$  value into Equation 3.

## Method

**Observers.** The observers were 24 paid University of Washington students with normal or corrected-to-normal vision.

**Apparatus.** The experiment was executed in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The computer was a Macintosh G4 driving two Apple 17-in. Studio Display monitors. One of the monitors (the *near monitor*), along with the keyboard, was placed at a normal viewing distance—that is, approximately, 1.5 ft from the observer. It was used to display the filtered pictures. The other monitor (the *far monitor*) was placed 8 ft away and was used to display the pictures that varied in size. The resolution of both monitors was set to  $1,600 \times 1,200$  pixels. The far monitor's distance from the observer was set so as to address the resolution problem that we raised earlier—that is, to allow shrinkage of a picture without concomitantly lowering the effective resolution of the display medium: From the observer's perspective, the far monitor screen had a pixel density of approximately 224 pixels per degree of visual angle, thereby rendering it unlikely that the pixel reduction associated with shrinking would be perceptible to the observer.

**Materials.** Four faces, two male and two female, created with the FACES Identikit program, were used as stimuli. They are shown in Figure 6. Each face was rendered as a  $1,100 \times 900$  pixel grayscale image, luminance-scaled so that the grayscale values ranged from 0 to 255 across pixels. For each of the four faces, six test images were created. They were sized so that, when presented on the far monitor, their visual angles would, to the observer seated at the near monitor, equal the visual angles subtended by real faces seen from six test distances: 21, 36, 63, 108, 185, and 318 ft.

**Design and Procedure.** Each observer participated in eight consecutive *sessions*, each session involving one combination of the 4 faces  $\times$  2 filter classes ( $r = 3$  and  $r = 10$ ). The 8 face  $\times$  filter class combinations occurred in random order but were counterbalanced so that, over the 24 observers, each combination occurred exactly three times in each of the eight sessions. Each session consisted of 10 *replications*. A replication consisted of six *trials*, each trial involving one of the six test distances. Within each replication, the order of the six test distances was randomized.

On each trial, the following sequence of events occurred. First, the test picture was presented on the far screen, where it remained unchanged throughout the trial. Simultaneously, a randomly selected 1 of the 41 comparison faces appeared on the near screen. The ob-



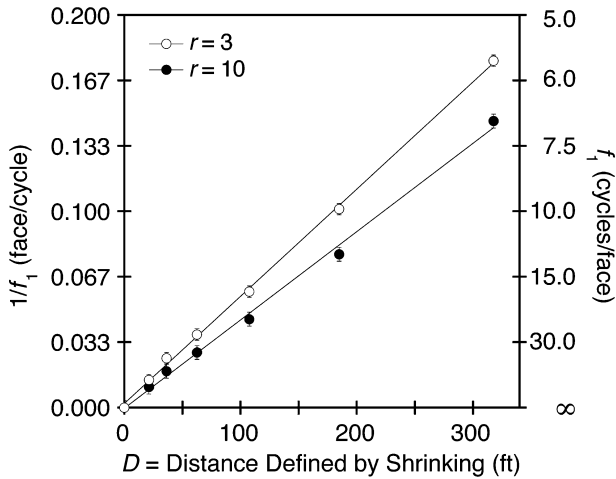
Figure 6. Faces used in Experiment 1.

server was then permitted to move freely back and forth through the sequence of comparison faces, all on the near screen. This was done using the left and right arrow keys: Pressing the left arrow key caused the existing comparison image to be replaced by the next blurrier image, whereas pressing the right arrow key produced the next clearer image. All of the 41 comparison faces were held in the computer's RAM and could be moved very quickly in and out of video RAM, which meant that moving back and forth through the comparison faces could be done very quickly. If either arrow key was pressed and held, the comparison image appeared to blur or deblur continuously, and to transit through the entire sequence of 41 comparison faces from blurriest to clearest or vice versa took approximately a second. Thus, the observer could carry out the comparison process easily, rapidly, and efficiently.

As was noted earlier, the observer's task was to select the comparison stimulus that best matched the perceived informational content of the test stimulus. In particular, the following instructions were provided: "On each trial, you will see what we call a *distant face* on the far monitor (indicate). On the near monitor, you will have available a set of versions of that face that are all the same size (large) but which range in clarity. We call these *comparison faces*. We would like you to select the one comparison face that you think best matches how well you are able see the distant face." The observer was provided unlimited time to do this on each trial, could roam freely among the comparison faces, and eventually indicated his or her selection of a matching comparison face by pressing the up arrow. The response recorded on each trial was the cutoff frequency,  $f_1$  in cycles/face, of the filter used to generated the matching comparison face.

## Results

Recall that the prediction of the distance-as-filtering hypothesis is that  $1/f_1$  is proportional to size-defined distance,  $D$  (see Equation 6). Our first goal was to test this



**Figure 7. Experiment 1 results: Reciprocal of measured  $f_1$  values (left-hand ordinate) plotted against distance,  $D$ . Right-hand ordinate shows  $f_1$  values. Solid lines are best linear fits. Each data point is based on 960 observations. Error bars are standard errors.**

prediction for each of our filter classes,  $r = 3$  and  $r = 10$ . To do so, we calculated the mean value of  $1/f_1$  across the 24 observers and four faces for each value of  $D$ . Figure 7 shows mean  $1/f_1$  as a function of  $D$  for both  $r$  values, along with best-fitting linear functions. It is clear that the curves for both  $r$  values are approximated well by zero-intercept linear functions—that is, by proportional relations. Thus, the prediction of the distance-as-filtering hypothesis is confirmed.

Given confirmation of the proportionality prediction, our next step is to estimate the human MTF for each value of  $r$ . To estimate the MTF, it is sufficient to estimate the cutoff value,  $F_1$ , which is accomplished by

computing the best-fitting zero-intercept slope<sup>2</sup> of each of the two Figure 7 functions, equating each of the slope values to  $1/(43F_1)$ , the predicted proportionality constant, and solving for  $F_1$  (again, see Equation 6). These values are 52 and 42 cycles/deg for measurements taken from stimuli generated by the  $r = 10$  and  $r = 3$  filters, respectively. Note that the corresponding estimates of the rolloff frequency  $F_0$ —the spatial frequency at which the MTF begins to descend from 1.0 en route to reaching zero at  $F_1$ —are approximately 5 and 14 cycles/deg for the  $r = 10$  and  $r = 3$  filters.

Is there any basis for distinguishing which of the two filters is better as a representation of the human MTF? As is suggested by Figure 7, we observed the  $r = 3$  filter to provide a slightly higher  $r^2$  value than did the  $r = 10$  filter (.998 vs. .995). Table 1 provides additional  $F_1$  statistics for the two filter classes. Table 1, column 2 shows the mean estimated  $F_1$  for each of the four individual faces, obtained from data averaged across the 24 observers. Columns 3–6 show data based on estimating  $F_1$  for each individual observer–face–filter combination and calculating statistics across the 24 observers. The data in Table 1 indicate, in two ways, that the  $r = 3$  filter is more stable than the  $r = 10$  filter. First, the rows marked “ $SD$ ” in columns 2–4 indicate that there is less variability across faces for the  $r = 3$  filter than for the  $r = 10$  filter. This is true for  $F_1$  based on mean data (column 2) and for the median and mean of the  $F_1$  values across the individual faces (columns 3 and 4). Second, columns 5 and 6 show that there is similarly less variability across the 24 observers for the  $r = 3$  than for the  $r = 10$  filter. Finally, the difference between the median and the mean was smaller for the  $r = 3$  than for the  $r = 10$  filter, indicating a more symmetrical distribution for the former.

**Table 1**  
**Filter Cutoff Frequencies ( $F_1$ ) Statistics for Two Filter Classes**  
**( $r = 10$  and  $r = 3$ ) and the Four Faces (Two males, M1 and M2, and**  
**Two females, F1 and F2)**

Face	Mean Based on Averaged Data	Computed Across Observers			
		Median	Mean	$SD$	Range
<i>r = 10 Filter</i>					
M1	51.0	50.5	60.0	27.1	119
M2	54.7	61.9	60.7	19.5	74
F1	53.1	53.3	57.2	17.6	70
F2	48.0	51.8	54.6	18.1	79
<i>M</i>	51.7	54.4	58.1	20.6	85.5
<i>SD</i>	2.5	4.5	2.4		
<i>r = 3 Filter</i>					
M1	44.3	44.9	46.8	14.0	53
M2	45.7	46.1	46.0	11.0	46
F1	41.3	43.1	42.7	10.6	43
F2	40.5	40.1	42.3	9.8	34
<i>M</i>	42.9	43.6	44.5	11.3	44.0
<i>SD</i>	2.1	2.3	2.0		

Note—In the rows labeled *M* and *SD*, each cell contains the mean or standard deviations of the four numbers immediately above it; that is, they refer to statistics over the four faces for each filter class. The columns labeled Mean, *SD*, and Range refer to statistics computed over the 24 observers.

Given that stability across observers and materials, along with distributional symmetry, indicates superiority, these data indicate that the  $r = 3$  filter is the better candidate for representing the MTF in this situation. Our best estimate of the human MTF for our Experiment 1 matching task is, therefore,  $F_0 = 14$ , and  $F_1 = 42$ . We note that these are not the values used to create Figure 4; the filter used for the demonstrations in Figure 4 was derived from Experiments 2–4, in which face recognition was used.

## Discussion

The goal of Experiment 1 was to determine the viability of the distance-as-filtering hypothesis. By this hypothesis, the deleterious effect of distance on face perception is entirely mediated by the loss of progressively lower image frequencies, measured in cycles/face, as the distance between the face and the observer increases. The prediction of this hypothesis is that, given the correct filter corresponding to the human MTF in a face perception situation, a face shrunk to correspond to a particular distance  $D$  is spatially filtered, from the visual system's perspective, in a way that is entirely predictable. This means that if a large-image face is filtered in exactly the same manner, it should be matched by an observer to the shrunken face—that is, the observer should conclude that the shrunken face and the appropriately filtered large face look alike, in the sense of containing the same spatial information.

Because there are not sufficient data in the existing literature, the correct MTF is not known, and this strong prediction could not be tested. What we did, instead, was to postulate two candidate MTFs and then estimate their parameters from the data. The general prediction was that the function relating the reciprocal of measured filter cutoff frequency,  $1/f_1$ , should be proportional to  $D$ , distance defined by size. This prediction was confirmed for both candidate MTFs, as indicated by the two functions shown in Figure 7. Although the gross fits to the data were roughly the same for the two candidate MTFs (they were both excellent), a criterion of consistency over observers and stimuli weighed in favor of the  $r = 3$  filter over the  $r = 10$  filter. We reiterate that our estimated  $r = 3$  filter, in combination with an  $F_1$  value of 42 cycles/deg, implies an MTF that passes spatial frequencies perfectly up to 14 cycles/deg and then falls, reaching zero at 42 cycles/deg.

How does this estimate comport with past data? Georgeson and Sullivan's (1975) data, discussed earlier, indicated a fairly flat human suprathreshold MTF up to at least 25 cycles/deg, which is certainly higher than the 14 cycles/deg rolloff frequency that we estimated here. Georgeson and Sullivan used a different task from the present one: Their observers adjusted contrast of sine wave gratings of specific spatial frequencies to match a 5 cycles/deg standard, whereas the present observers matched spatial frequency composition of constant contrast faces to match different-sized test stimuli. Because their highest

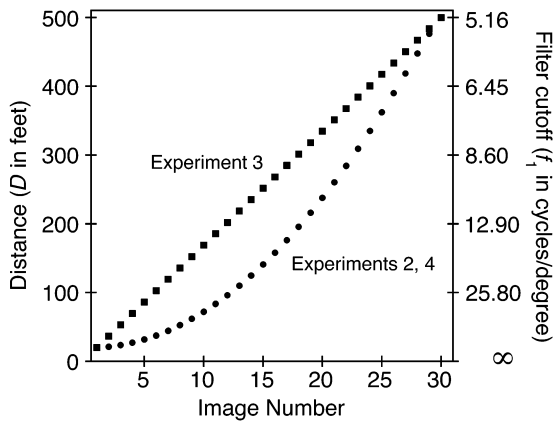
comparison grating was 25 cycles/deg, we do not know whether and how the function would have behaved at higher spatial frequencies. Similarly, the present measurements of MTF shape are limited because we used  $r$  values of only 3 and 10. It is possible that if, for instance, we had used an  $r = 2$  filter, we would have estimated a larger rolloff frequency.

Our estimated MTF shape does accord well with the data described earlier, reported by Hayes et al. (1986). They reported an experiment in which faces were (1) filtered with band-pass filters centered at difference image frequencies (expressed in terms of cycles/face) and (2) presented at different viewing distances, which, for a given image frequency distribution, affected the distribution of absolute spatial frequencies (expressed in cycles/degree). As we articulated earlier, Hayes et al.'s data can be nicely accommodated with the assumptions that (1) the human MTF is low-pass for their experiment and (2) it passed spatial frequencies perfectly up to somewhere between 10 and 20 cycles/deg.

In Experiments 2–4, we explored whether the MTF filter estimated from the matching task used in Experiment 1 is also appropriate for face identification. To anticipate, the filter that we estimated in Experiments 2–4 had smaller values of  $F_0$  and  $F_1$ ; that is, it is scaled toward lower spatial frequencies. In our General Discussion section, we will consider why that may happen. For the moment, we note that one common observer strategy in Experiment 1 was to focus on small details (e.g., a lock of hair), judge how visible that detail was in the distant face, and then adjust the comparison face so that it was equally visible. This focusing on small details may mean that the observers in Experiment 1 viewed their task as more like looking at an eye chart than at a face and that whatever processes are *special* to face processing were minimized in Experiment 1. In Experiments 2–4, we dealt with the same general issues as those in Experiment 1, but using a task that is unique to face processing: recognition of known celebrities.

## Experiment 2 Celebrity Recognition at a Distance: Priming

In Experiments 2–4, the same logic was used as in Experiment 1, relating distance defined by low-pass spatial filtering to distance defined by size. However, the stimuli, task, and dependent variables were all very different. On each of a series of trials in Experiments 2–4, the observers identified a picture of a celebrity. The picture began either very small or very blurry—so small or blurry as to be entirely unrecognizable—and then gradually became clearer by either increasing in size or deblurring. The observer's mission was to identify the celebrity as early as possible in this clarification process, and the clarity point at which the celebrity was correctly identified was recorded on each trial. This clarity point was characterized as the distance implied by size,  $D$ , in the case of increasing size, and as filter cutoff frequency in cycles/face,  $f_1$ , in the case of deblurring. As in Experiment 1, the cen-



**Figure 8. Experiments 2–4: degree of degradation plotted against image number. Higher image numbers represent more degraded pictures—that is, pictures that are shrunk to represent greater distances or are filtered with a lower cutoff frequency. The left ordinate represents virtual distance,  $D$ , whereas the right ordinate represents filter cutoff frequency,  $f_1$ . Note that decreasing  $f_1$  corresponds to increasing blur.**

tral prediction of the distance-as-filtering hypothesis is that  $1/f_1$  is proportional to  $D$  with a proportionality constant of  $1/(43F_1)$ . Again as in Experiment 1, given confirmation of this prediction,  $F_1$  can be estimated by equating the observed constant of proportionality to  $1/(43F_1)$  and solving for  $F_1$ .

Given confirmation of the prediction, it becomes of interest to determine how robust the estimate of the human MTF is. We evaluated such robustness in two ways. The first was to compare the MTF estimates based on recognition (Experiments 2–4) with the estimate based on matching (Experiment 1). The second was to compare MTF estimates within each of Experiments 2–4 under different circumstances. In particular, in each of Experiments 2–4, we implemented a dichotomous independent variable that, it was assumed on the basis of past data, would affect identification performance. The variable was *cognitive*, or *top-down*, in that variation in it did not affect any physical aspect of the stimulus; rather, it affected only the observer's expectations or cognitive strategies. The purpose of incorporating these variables was to test the *substitutability* of the presumed human MTF (Palmer, 1986a, 1986b; see also Bamber, 1979). The general idea of substitutability is that some perceptual representation, once formed, is unaffected by variation in other factors—that is, substitutability implies a kind of independence. The distance-as-filtering hypothesis is that filtering a face by some specified amount produces a perceptual representation that is functionally equivalent to the representation that is obtained when the face is viewed from a distance. Addition of the stronger substitutability hypothesis is that this equivalence is unaffected by changes in other, nonperceptual variables. Therefore, the prediction of substitutability is that the

same MTF—that is, the same estimated value of  $F_1$ —should describe the relation between distance defined by size and distance defined by filtering for both levels of the cognitive variable.

In Experiment 2, this cognitive variable was *priming*: Half of the to-be-identified celebrities had been primed, in that the observer had seen their names a few minutes prior to the identification part of the experiment. The other half of the to-be-identified celebrities had not been primed. Numerous past studies (e.g., Reinitz & Alexander, 1996; Tulving & Schacter, 1990) have indicated that such priming improves eventual recognition of the primed stimuli.

## Method

**Observers.** The observers were 24 paid University of Washington graduates and undergraduates with normal, or corrected-to-normal vision. All were raised in the United States and professed reasonable familiarity with celebrities, broadly defined.

**Apparatus.** The apparatus was the same as that in Experiment 1.

**Materials.** Pictures of 64 celebrities were obtained from various sources—principally, the Internet and glossy magazines. The celebrities came from all walks of celebrity life. They included actors, musicians, politicians, business figures, and sports figures, chosen so as to be recognizable by as many potential observers as possible. There were 43 males and 21 females. They were all rendered initially as grayscale images. They were scaled to be 500 pixels high but varied in width. Each image was luminance scaled so as to range in grayscale level from 0 to 255 across pixels.

Beginning with a single original photograph of each celebrity, two 30-image sets were created. Images in the first set varied in size so that, when shown on the far monitor, they subtended visual angles corresponding to distances ranging from 20 to 500 ft. Images in the second set were filtered using the  $r = 3$  filter class described in conjunction with Experiment 1. The cutoff values,  $f_1$ , ranged from 129 to 5.2 cycles/face. The exact function relating degree of degradation to degradation image number (1–30) was somewhat complex and represented an attempt to make the transition from one degraded image to the next less degraded image as perceptually similar as possible across all 30 degraded images for both the size increase and the deblurring procedures. The function is shown by the circles in Figure 8. The left ordinate refers to degradation defined in terms of size reduction (distance  $D$ ), whereas the right ordinate refers to degradation defined in terms of filtering (cutoff frequency  $1/f_1$ ; note the reciprocal scale). The sizes of the filtered images were the same as the sizes of the original images (500 pixels high  $\times$  variable widths).

**Design and Counterbalancing.** The observers were run individually. Each observer participated in four blocks of 16 trials/block. Each trial involved a single celebrity whose image became clearer by either increasing in size or deblurring. All trials within a block involved just one clarification type, increasing in size or deblurring. For half the observers, the blur (B) and size (S) block sequence was BSSB, whereas for the other half, the sequence was SBBS. Within each block, half of the celebrities had been primed in a manner to be described shortly, whereas the other half of the celebrities had not been primed.

Each observer had a *mirror image* counterpart whose trial sequence was identical except that primed–unprimed was reversed. Therefore, 4 observers formed a complete counterbalancing module, and the 24 observers made up six such modules. The order of the 64 celebrities across trials was constant across the 4 observers within each counterbalancing module but was randomized anew for each new module.

**Procedure.** The observer was told at the outset of the experiment that on each of a number of trials, he or she would see a picture of a celebrity that, while initially degraded so as to be unrecognizable, would gradually become clearer to the point that it was entirely clear.

The priming manipulation was implemented as follows. At the start of each block, the observer was told, "In the next block, you will see 8 of the following 16 celebrities," and then saw 16 celebrity names, one at a time. The observer was instructed to try to form an image of each celebrity when the celebrity's name appeared and then to rate on a 3-point scale whether they thought that they would recognize the celebrity named if his or her picture were to appear. The sole purpose of the ratings was to get the observer to visualize the celebrities during the priming stage, and the ratings were not subsequently analyzed. As promised, 8 of the 16 celebrities whose pictures the observer subsequently viewed during the block had been named during the initial priming phase. The observers were assured, truthfully, that the 8 celebrity names seen during the priming stage that *did not* correspond to celebrities viewed during that block were of celebrities who would not be viewed at any time during the experiment.

Following instructions at the outset of the experiment, the observer was provided four blocks of two practice trials per block in the same block order, BSSB or SBBS, that he or she would encounter in the experiment proper. Each of the four practice blocks was the same as a block in the experiment proper, except that there were only two trials, preceded by two primed names, in each practice block. No celebrity named or shown at practice appeared in any form during the experiment proper.

Increasing-size images were shown on the far screen, while deblurring images were shown on the near screen. On each trial, the images became clearer at the rate of 500 msec/image. The observers were told that they were to watch the clarifying celebrity and make a guess about who it was whenever they believed that they could recognize the person. The observers responded by first pressing the space bar, which stopped the clarification process, and then typing in their guess. They were allowed to use any kind of information they wished about a celebrity to indicate that they recognized him or her. For instance, if the celebrity was Jennifer Lopez, a valid response would be "Jennifer Lopez" or "J-Lo" or "that actress who was in *Maid in Manhattan*" or "affianced to Ben Affleck," or anything else that identified Ms. Lopez as the depicted celebrity. Once a guess had been made, the clarifying process resumed. The observers were allowed to change their minds and make additional guesses if and when they chose. When the image reached the point of complete clarification, one of two questions appeared on the screen. If the observer had made at least one guess, as usually happened, the question was, "Do you still think that this is XYZ," where XYZ was the most recent guess. If the observer answered "y," the trial ended. If the observer answered anything else, the computer asked, "OK, who do you think it is," and the observer typed in his or her final guess. If, as occasionally happened, the observer had made no guess during the clarification process, the computer asked, "OK, who do you think it is," and the observer was required to guess.

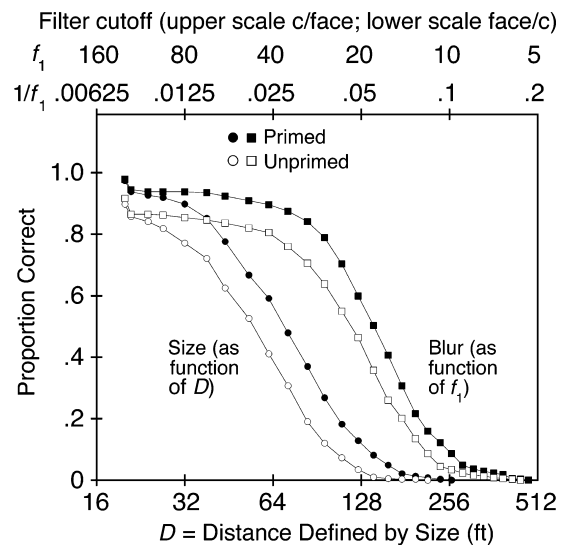
When the observer had completed the 64 trials, an initial analysis program was run. This program sequenced through all 64 trials. For each trial, it displayed the true name of the celebrity for that trial and then proceeded through all the responses that the observer had made on that trial (typically, only one response per trial, but occasionally more than one). The experimenter indicated whether each response was correct or incorrect by typing "y" or "n." This procedure was carried out with the observer still present, so the observer could provide occasional assistance in cases of ambiguous answers, bizarre spellings, idiosyncratic shorthand, and the like. Thus, for each trial was recorded the distance,  $D$  (in the case of increasing size), or the cutoff spatial frequency,  $f_1$  (in the case of decreasing blur) at which the celebrity was first correctly identified.

## Results

For each observer, the geometric mean point of initial identification was calculated for both primed and unprimed conditions. For the increasing-size conditions, this point was measured in distance,  $D$ , whereas for the deblurring conditions, it was measured in filter cutoff point,  $f_1$ . Note that for distance, larger indicates better performance: A longer distance,  $D$ , means that the observer was able to identify the celebrity at a more degraded point. For blurring, smaller is better: A smaller cutoff frequency,  $f_1$ , means that the observer was able to identify the celebrity at a more degraded point.

The priming manipulation had the expected effect: Primed celebrities could be recognized at a point that was  $24.5\% \pm 6.4\%$  smaller and  $22.8\% \pm 11.8\%$  more blurred than unprimed celebrities.<sup>3</sup>

For increasing-size trials, the proportion of correctly recognized celebrities was calculated for each of the 30 distances,  $D$ , defined by size. Likewise, for deblurring trials, the proportion correct for celebrities recognized was calculated for each of the 30 cutoff frequencies,  $f_1$  in cycles/face. The results, depicted as psychophysical functions, are shown in Figure 9. Note that in Figure 9, there are three separate abscissas, all scaled logarithmically. The bottom abscissa shows  $D$  and is relevant to the increasing-size trials. The two top abscissas are relevant



**Figure 9. Experiment 2 data: Proportion of celebrities recognized as a function of degradation level.** The two left curves show data for degradation by virtual distance (size reduction) and are plotted as a function of distance,  $D$ , on the bottom abscissa. The two right curves show data for degradation by filtering and are plotted as a function of cutoff frequency,  $f_1$ . The prediction of the distance-as-filtering hypothesis is that the two primed curves (closed symbols) and the two unprimed curves (open symbols) are both horizontally parallel; modulation transfer function substitutability further implies that primed and unprimed curves be separated by the same amount. Each data point is based on 384 observations.

to the decreasing-blur trials. They show  $f_1$  (for intuitive clarity), along with  $1/f_1$  (for ease of the mathematical manipulation to be carried out below). The two left curves (circle curve symbols) correspond to increasing size, whereas the two right curves (square curve symbols) correspond to decreasing blur. The open-symbol curves correspond to unprimed curves, whereas the closed-symbol curves correspond to primed curves.

The four curves are smooth and regular, with one exception: There is a slight jump in performance at the least degraded (largest or least blurry) image for all four blur/size  $\times$  priming conditions. This is almost certainly an artifact that stems from the observers' being forced to guess following presentation of the least degraded image if they had not already made a response. These four artifactual points are not included in the analysis to be described next.

To assess the prediction of the distance-as-filtering hypothesis, we begin with Equation 5, which relates distance defined by size ( $D$ ) to distance defined by filtering ( $f_1$ ). Performance in the increasing-size condition is predicted to equal performance in the deblurring condition when

$$p\left(\frac{43F_1}{f_1}\right) = p(D), \quad (7)$$

where  $p(x)$  refers to the proportion of celebrities recognized at degradation level  $x$ . Changing to logarithms and rearranging terms,

$$p\left[\log(43F_1) + \log\left(\frac{1}{f_1}\right)\right] = p[\log(D)]. \quad (8)$$

Thus, by Equation 8, the prediction of the distance-as-filtering hypothesis is that the curves corresponding to deblurring [measured in units of  $\log(1/f_1)$ ] and increasing size [measured in units of  $\log(D)$ ] will be horizontally parallel, separated by a constant equal to  $\log(43F_1)$ .

Given this foundation, we can identify four possible outcomes that would imply four corresponding conclusions. They are, in increasing degree of conclusion strength, as follows.

1. Failure of Equation 8. The deblurring and distance curves are not described by Equation 8 (i.e., they are not log horizontally parallel). This would imply disconfirmation of the distance-as-filtering hypothesis.

2. Confirmation of Equation 8, but with different shift values for the two priming conditions. Both priming levels are adequately described by Equation 8, but with different shift magnitudes for the two levels. This would imply confirmation of the distance-as-filtering hypothesis but would also imply different MTFs for the two priming levels—that is, failure of substitutability.

3. Confirmation of Equation 8 with the same shift value for the two priming conditions. Both priming levels are adequately described by Equation 8, with the same shift magnitude for the two levels. This would imply confirmation of the distance-as-filtering hypothe-

sis and would imply that a single MTF is adequate to describe the two priming levels.

4. Confirmation of Equation 8 with the same shift value for the two priming conditions whose magnitude implies the  $F_1$  value of 42 cycles/deg that was estimated in Experiment 1. This would imply a single filter upon which (at least) two quite different visual perception tasks—matching and recognition—are based.

Our results imply Conclusion 3. Figure 10 shows the result of simultaneously shifting primed and unprimed size curves by 3.13 log units. For visual clarity, the two primed curves have been shifted to the right by an additional 0.5 log unit. The alignment of the size curves with their blur counterparts is almost perfect. The estimate of  $F_1$  corresponding to the observed 3.13 log-unit shift is  $31.5 \pm 2.7$  cycles/deg.

## Discussion

The data in Experiment 2 provide clear support for the distance-as-filtering hypothesis. The psychophysical functions for the size and filtering degradation techniques are log-parallel—that is, are proportional to one another.

Priming, although substantially affecting performance, did not affect the relation between distance defined by size and distance defined by blur. Thus, the nature of the assumed MTF is inferred to be unaffected by priming. This constitutes confirmation of the substitutability hypothesis that we described earlier. To foreshadow, we observed the same kind of confirmation in Experiments 3 and 4.

The estimated MTF had a cutoff frequency,  $F_1$ , of  $31.5 \pm 2.7$  cycles/deg. This is somewhat different from

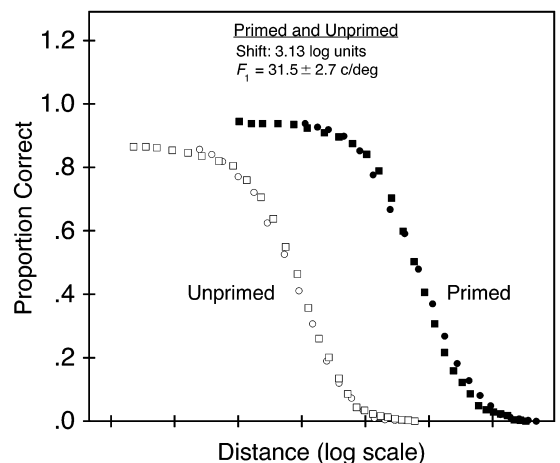


Figure 10. Experiment 2 data: Both the primed and the unprimed filtered curves from Figure 9 have been shifted by 3.13 log units. For visual clarity, the primed curves are arbitrarily shifted 0.5 log units to the right. For both the primed and the unprimed curves, circles represent increasing-size trials, whereas squares represent decreasing-blur trials.



the corresponding MTF estimated in Experiment 1, whose  $F_1$  was approximately 42 cycles/deg. This means that, within the context of the distance-as-filtering hypothesis, there is not a single MTF for all tasks involving face processing. We will return to these issues in our General Discussion section.

### Experiment 3 Celebrity Recognition at a Distance: Perceptual Interference

Experiment 3 was largely identical to Experiment 2, except that instead of priming, *degradation starting point* was varied. On half the trials, the celebrity's picture began its degradation process at a starting point termed *far* (think "celebrity far away"), which was identical to the starting point in Experiment 1. On the other half of the trials, the starting point, termed *near*, was half as degraded as the far starting point. Beginning with a well-known report by Bruner and Potter (1964), numerous investigators have demonstrated that in this kind of experiment, beginning the degradation process at a more degraded point decreases the observer's eventual ability to recognize what the picture depicts, as compared with beginning at a less degraded point. This finding is called the *perceptual interference effect*. Based on these past data, we anticipated (correctly) that performance would be better with a near than with a far starting point. The substitutability prediction is that the estimated correspondence between distance defined by filtering and distance defined by size should not differ for the two starting points.

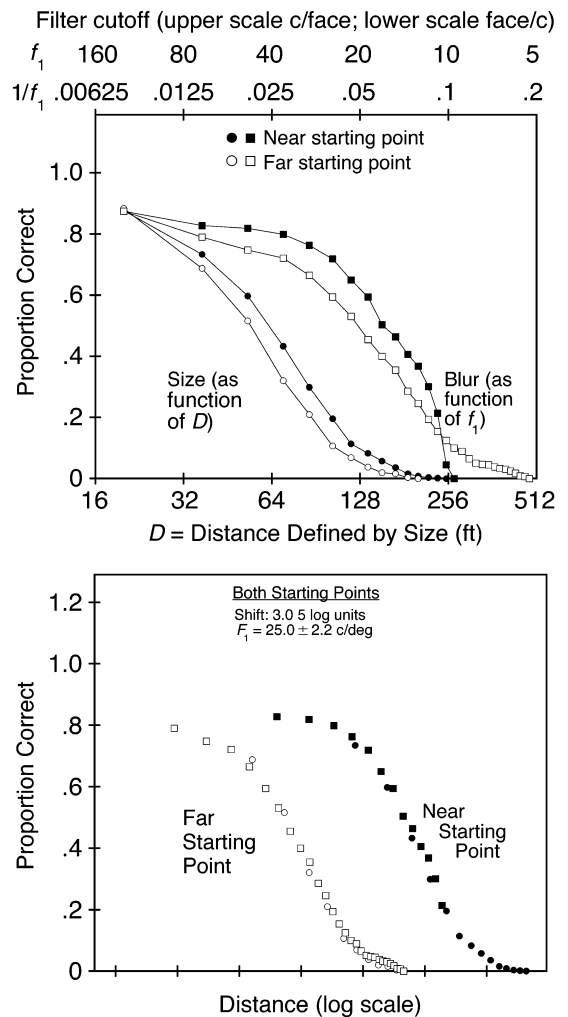
#### Method

**Observers.** The observers were 24 paid University of Washington graduates and undergraduates. All were raised in the United States and professed reasonable familiarity with celebrities, broadly defined.

**Apparatus.** The apparatus was the same as that used in Experiments 1–2.

**Materials.** The same celebrity pictures as those in Experiment 2 were used in Experiment 3. There were still 30 sized images and 30 corresponding filtered images for the varying degrees of degradation, with the same most degraded and least degraded points. However, the particular degradation levels that were used varied from least to most degraded in equal distance intervals for distance defined by size and, correspondingly, equal  $1/f_1$  intervals for distance defined by filtering as is shown by the diamonds in Figure 8. The two starting points, *far* and *near*, were images whose sizes corresponded to 500 and 250 ft for the distance-as-size images and images filtered with cutoff frequencies of 5.2 and 10.3 cycles/face for distance-as-filtering images.

**Design, Procedure, and Counterbalancing.** Design, procedure, and counterbalancing, were as in Experiment 2 except for the following. First, of course, there was no priming procedure beginning each block. Second, the primed/unprimed manipulation of Experiment 1 was replaced by the near/far-start manipulation just described. Third, on each trial, the images became clearer at the rate of 500 msec/image in the far-start condition and 1,000 msec/image in the near-start condition. Because there were only half as many images to go through in the near-start, as compared with the far-start, condition, this meant that total time from start to finish would, if uninterrupted, be the same in the near- and the far-start conditions.



**Figure 11. Experiment 3 data:** The two panels correspond to Figures 9 and 10. Upper panel shows raw data, and lower panel shows shifted data. Both starting point curves have been shifted by the same amount. For visual clarity, the near starting point curves are shifted 0.5 log units to the right in the lower panel. Each data point is based on 448 observations. In both panels, for both starting point curve pairs, circles represent increasing-size trials, whereas squares represent decreasing-blur trials.

#### Results and Discussion

The starting point manipulation had the expected effect: Celebrities could be recognized at a point that was  $20.7\% \pm 10.2\%$  smaller and  $27.3\% \pm 15.5\%$  more blurred in the near- than in the far-start condition.

Figure 11 shows the results in Experiment 3. The top panel, like Figure 9, shows the raw data. The “clearest position” artifact described in conjunction with Experiment 2 is again apparent. Moreover, there is a new artifact in Experiment 3 that is uninteresting but complicated. It is this. In all the conditions, the most degraded point at which a celebrity could possibly be identified was, of course, the starting point. The normal starting point—that

is, the *far* starting point in Experiment 3—is sufficiently degraded that no celebrity was ever recognized at that point by any observer. The near starting point in Experiment 3, though, is one at which celebrities *were* occasionally identified. Obviously, in the near starting point condition, no celebrity can be identified at any point that is more degraded than the near starting point. This limitation affected the most degraded two points of the 250 starting point blur condition, which, as is apparent in the top panel of Figure 11, are artificially low.

Figure 11, bottom panel, like Figure 10, shows the result of shifting the near and the far starting point curves by the same amount—in this case, 3.05 log units. Again, the artifactual points described above are excluded from these shifted data. The conclusions implied by the results in Experiment 3 are analogous to those in Experiment 2. First, the distance-as-filtering hypothesis is confirmed by the log-parallel distance and filtered curves. Second, a new top-down variable, starting distance, has no effect on the estimated filter. Finally, the MTF in Experiment 3, calculated as in Experiment 2 from the observed best shift, has a cutoff value,  $F_1$ , of  $25.0 \pm 2.2$  cycles/deg.

#### Experiment 4 Celebrity Recognition at a Distance: Hindsight Bias

Experiment 4 was largely identical to Experiments 2 and 3. Here, we varied what we term *normal/hindsight*. On half the blocks (the *normal* blocks), the celebrity's picture was shown in a normal fashion—that is, unprimed, with a long starting point. On the other half of the blocks (the *hindsight* blocks), the observer was shown a large, clear picture of the celebrity at the start of the trial. He or she then proceeded through the clarification process—increasing size or deblurring, as usual—but was asked to indicate at what point he or she *would have recognized* the celebrity, if they had not already known who it was. On the basis of past data (e.g., Bernstein, Atance, Loftus, & Meltzoff, 2004; Harley, Bernstein, & Loftus, 2002; Harley, Carlsen, & Loftus, 2004), we anticipated that the observer would claim the celebrity to be recognizable at a more degraded point in hindsight, as compared with the normal procedure. The substitutability prediction is that the estimated MTF should not differ for the normal/hindsight manipulation.

#### Method

**Observers.** The observers were 32 paid University of Washington graduates and undergraduates. All were raised in the United States and professed reasonable familiarity with celebrities, broadly defined.

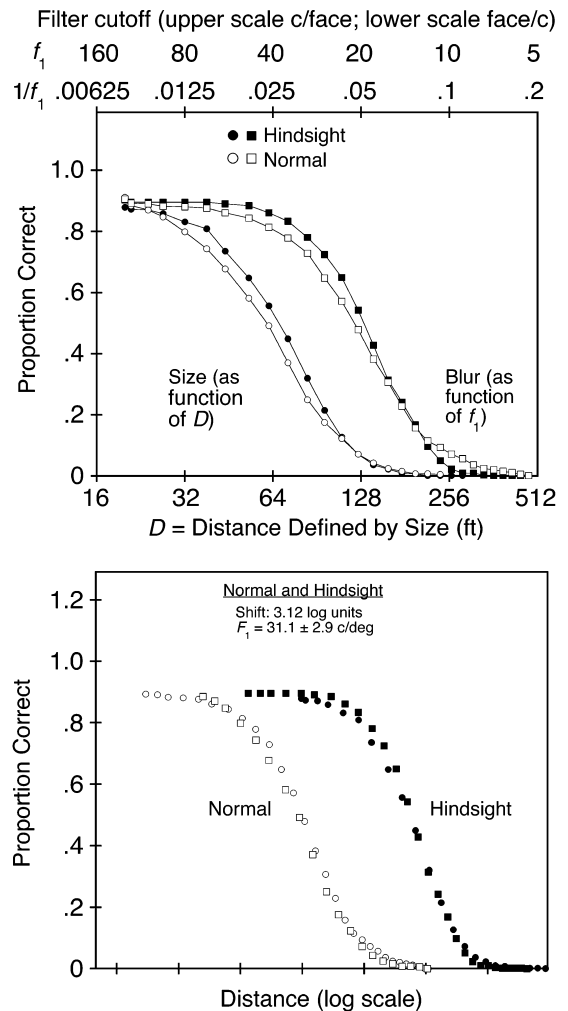
**Apparatus.** The apparatus was the same as that in Experiments 1–3.

**Materials.** A new set of 64 grayscale images of celebrity faces was used in Experiment 4 (33 were taken from Experiments 2 and 3, whereas 31 were new<sup>4</sup>). As in Experiment 2, there were 30 sized images and 30 corresponding filtered images created for each celebrity face. The same set of degradation levels as that in Experiment 2 was used in Experiment 4.

**Design, Procedure, and Counterbalancing.** In general, design, procedure, and counterbalancing were as in Experiments 2 and 3. Again, there were four 16-trial blocks, shown in a BSSB order for 16 observers and in SBBS order for the other 16 observers. Within each 16-trial block, a random 8 of the trials were normal (i.e., like the unprimed Experiment 2 trials), whereas the remaining 8 were hindsight trials. On each hindsight trial, the observer was, as indicated, shown a completely clear picture of the celebrity for that trial at the start of the trial but was instructed to ignore this knowledge in selecting the eventual identification point. Following presentation of the clear picture, the remainder of a hindsight trial was identical to a normal trial. Each observer had a mirror image observer whose trial sequence was identical, except that normal/hindsight trials were reversed.

#### Results and Discussion

The normal/hindsight manipulation had the expected effect, but it was small: Celebrities could be recognized



**Figure 12. Experiment 4 data:** The two panels are set up like those of Figure 11. In both panels, for both the normal and the hindsight curve pairs, circles represent increasing-size trials, whereas squares represent decreasing-blur trials. Each data point is based on 512 observations.

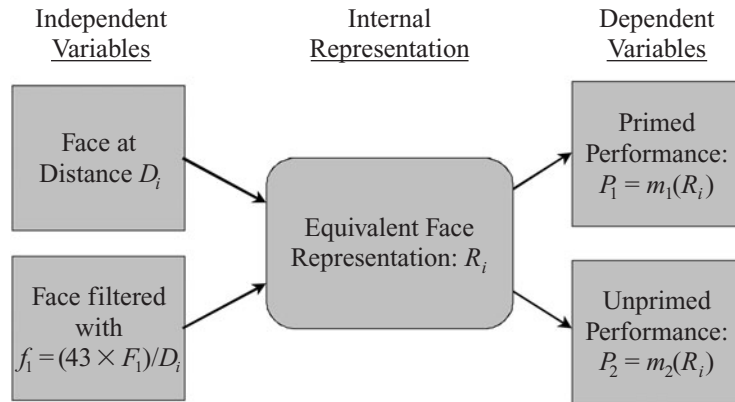


Figure 13. Depiction of distance-as-filtering hypothesis as applied to Experiments 1–4.

at a point that was  $11.2\% \pm 10.4\%$  smaller and  $6.0 \pm 10.6\%$  more blurred in the hindsight than in the normal condition.

Figure 12, like Figure 11, shows the results in Experiment 3: The top panel shows the main results, whereas the bottom panel shows the shifted curves. Again, the leftmost top-panel artifactual points are excluded from the shifted data. The conclusions implied by the results in Experiment 4 are analogous to those in Experiments 2 and 3. First, the distance-as-filtering hypothesis was confirmed by the log-parallel distance and filtered curves. Second, a third top-down variable, normal/hindsight, had no effect on the estimated filter (although, since the effect of normal/hindsight was quite small, this result is not as meaningful as it was in Experiment 2 or 3). Finally, the MTF in Experiment 4, calculated as in Experiments 2 and 3 from the observed best shift, has a cutoff value,  $F_1$ , of  $31.1 \pm 2.9$  cycles/deg.

## GENERAL DISCUSSION

For Experiments 1–4, two different tasks have been reported: matching face blurriness to face size and identifying celebrities. In each task, the question was, What is the relation between the filter (characterized by  $f_1$ ) and distance (defined by  $D$ ) that gives rise to equal performance? For both tasks, to a high degree of precision, the answer was that  $D$  and  $f_1$  are inversely proportional to one another—that is, Equation 5 was confirmed.

Within the context of the distance-as-filtering hypothesis, this finding allows estimation of the parameters of the hypothesized filter corresponding to the human MTF. Between tasks, the MTF parameters were somewhat different: The estimates of  $F_1$  were 42 cycles/deg for the matching task and 25–31 cycles/deg for the celebrity recognition task. For all three celebrity recognition experiments, however, the substitutability prediction was confirmed: The estimated MTF was the same—that is, the  $F_1$  estimate was the same—for each of two

priming levels in Experiment 2, for each of two starting point levels in Experiment 3, and for both normal and hindsight conditions in Experiment 3. This implies a certain degree of robustness in the MTF estimates.

## Face Degradation and Face Representation

Seeing a face, or a picture of a face, produces a representation of the face that can be used to carry out various tasks. The greater the distance at which the face is viewed, the poorer is the face's representation. The distance-as-filtering hypothesis asserts that the representation of the face at a particular distance,  $D$ , can be obtained by either shrinking the face or low-pass filtering it to simulate the effect of distance on the image frequency spectrum.

Figure 13 depicts the hypothesis and its application to the experiments that we have reported. The logic of Figure 13 is this. The left boxes represent distance,  $D$ , and filtering,  $f_1$ . Note that having specified the form of the MTF (Equation 4) and the parameter  $r$  (which we assume to equal 3 for this discussion), a particular filtered face is completely specified by  $f_1$ , the filter's cutoff frequency in cycles/face. The expression  $f_1 = (43 \times F_1)/D$  in the lower left box, along with the middle "Internal Representation" box, signifies that when  $f_1$  is proportional to  $1/D$  with a proportionality constant of  $43F_1$ , the representations of the face are functionally equivalent.

This foundation is sufficient to account for the data in Experiment 1: If one face, shrunk to correspond to distance,  $D$  and another face filtered by  $f_1 = (43 \times F_1)/D$  produce equivalent representations, they will match.

To account for the data in Experiments 2–4, we must specify how celebrity identification performance is related to the representation. We will not attempt to do this in detail here, but rather, we will sketch two theoretical strategies. The first is to assume that the representation,  $R$ , can be formulated as a single number on a unidimensional scale (see Loftus & Busey, 1992, for discussions of how this might be done) with better representations cor-

responding to higher values of  $R$ . Performance would then be a simple monotonic function,  $m$ , of  $R$ . The second strategy is to assume that  $R$  is multidimensional (e.g., O’Toole, Wenger, & Townsend, 2001; Townsend, Solomon, & Smith, 2001), with each dimension being monotonically related to the representation’s quality. Performance would again be a function,  $m$ , of  $R$  that is monotonic in all arguments.

In either case, there would be two such monotonic functions corresponding to the two levels of each of the cognitive independent variables in Experiments 2–4 (we arbitrarily use primed/unprimed as our example in Figure 13). Thus if  $m_1$  were defined to be the function for the “difficult” condition, it would correspond to unprimed (Experiment 2), far starting point (Experiment 3), or normal (Experiment 4), whereas  $m_2$  would correspond to primed, near starting point, or hindsight. Thus, with a given representation  $R$ ,  $m_2$  would produce higher performance than would  $m_1$ . An important property of the data captured by the Figure 13 representation, however, is that in all cases the *cognitive* independent variable—priming, starting point, and normal/ hindsight—does not affect the nature of the MTF as defined by  $F_1$ . This is Figure 13’s reflection of the substitutability hypothesis.

### Different MTFs for Different Tasks

The simplest incarnation of the distance-as-filtering hypothesis would be that, under comparable physical circumstances, the human visual system acts as a fixed spatial filter. A comparison of Experiment 1, on the one hand, and Experiments 2–4, on the other, weighs against this possibility. For two tasks (matching and identification) the estimated MTFs differed: For Experiment 1, the estimated MTF had a cutoff of  $F_1 = 42$  cycles/deg, whereas for Experiments 2–4, the estimates of  $F_1$  were approximately 31, 25, and 31 cycles/deg.

It is possible that the difference in estimated MTFs results from the different stimuli used for the two tasks; the four computer-generated faces shown in Figure 6 were used in the matching task, whereas celebrity photos were used for the identification task. However, the two kinds of faces are not very different physically, and there is no a priori reason to anticipate that one, as compared with the other, would produce a different MTF estimate. So it seems unlikely that the difference in the estimated MTF that emerged resulted entirely from the stimulus set difference.

It *does* seem plausible, however, that the different *tasks* might produce the different MTF estimates. Schyns and his colleagues have, in recent years, made a case for what they refer to as “flexible scale usage,” which refers to the proposition that people use spatial scales (e.g., as instantiated in different image frequency bands) to varying degrees and in different orders, depending on the task at hand (Gosselin & Schyns, 2001; Oliva & Schyns, 1997; Schyns & Oliva, 1999; see Morrison & Schyns, 2001, for a review). Thus, for example, Schyns and Oliva showed observers hybrid faces. A hybrid face is a “double exposure” of two superimposed faces, one composed

of only low spatial frequencies (below 8 cycles/face) and the other composed only of high spatial frequencies (above 24 cycles/face). The two faces differed on three dimensions: male/female, expressive/nonexpressive, and angry/happy. The observers were asked to categorize the faces along one of these dimensions after the face had been briefly presented, and the selected member of the hybrid—the low spatial frequency member or the high spatial frequency member—was noted on each trial. The investigators found that different categorization tasks influenced which member of the pair was chosen: In their Experiment 1, for instance, an observer performing an expression/no-expression categorization chose the low spatial frequency member of the hybrid 38% of the time, whereas an observer performing the happy–angry categorization chose the low spatial frequency member of the exact same hybrid 66% of the time. The implication, therefore, was that the observers’ perception of what they were seeing—and in particular, which image spatial frequency band dominated their perception—was strongly influenced by the task that they were carrying out.

75%:  $f_1 = 39$  c/face,  $D = 34$  ft



25%:  $f_1 = 14$  c/face,  $D = 77$  ft



**Figure 14.** Low-pass filtered stimuli corresponding to 75% correct (top) and 25% correct (bottom). c/face, cycles per face.

The differences between the tasks in Experiment 1, on the one hand, and those in Experiments 2–4, on the other, likely biased the observers to attend to different spatial scales. As we have already noted, the observers performing the matching task in Experiment 1 spontaneously reported a common strategy of focusing on small details of the (shrunk) test picture—for example, on a loose lock of hair—and then adjusting the blur of the comparison picture so as to mimic how well they could see it. This is a strategy that emphasizes high spatial frequency information and, indeed, does not even require that the observers be looking at a face: It seems likely that much the same results would have been obtained had the observers been looking at any complex visual stimulus. In contrast, the identification task in Experiments 2–4 required that the stimuli be treated as faces.

Despite these observations, however, our finding of different MTFs for different tasks, although analogous to Schyns's findings, goes beyond the observation that observers are attending to different image frequencies for different tasks. Whereas the data reported by Schyns and his colleagues imply attention to different image scales expressed in terms of cycles/face, our findings indicate a difference in something more fundamental—namely, a different absolute MTF for different tasks expressed in terms of cycles/degree. We do not have a ready explanation for why this happens; we can only speculate that it may be related to the alleged “specialness” of face processing (e.g., Farah, Wilson, Drain, & Tanaka, 1998), which, although clearly needed in Experiments 2–4, was not needed in Experiment 1.

### Recognizing Faces at a Distance

One of the goals of this research was to provide some quantitative information about the distance at which people can be recognized. What can we infer from our data?

**Celebrity identification.** Celebrity identification provides a good paradigm for investigating the ability to identify people who are known to the observer. Our data provide some estimates of the degree of clarity required for such identification. In particular, we estimated the degree of clarity required to identify the celebrities at 25% and 75% rates for three comparable conditions in three identification experiments: unprimed in Experiment 2, far start in Experiment 3, and “normal” in Experiment 4. Averaged across the three experiments, the 25% and 75% identification level distances,  $D$ , were 77 and 34 ft, and the filters, expressed as cutoff,  $f_1$ , were 14 and 39 cycles/face. To provide a sense of the degree of degradation that these numbers represent, Figure 14 shows Julia Roberts, filtered so as to represent her presumed appearance at these two degradation levels.

A number of studies have reported data concerning spatial frequency bands that are optimal for face identification. Morrison and Schyns (2001) have summarized these studies. The common finding is that there is an optimal spatial frequency band centered around 11–20 cy-

cles/face height. It is not entirely clear, however, how these findings should be compared with ours, since there are a number of important differences in the experimental procedures.

First, these experiments did not involve identification of previously familiar individuals. Instead, one of two procedures was used: Observers were initially taught *names* corresponding to a small set of faces and then attempted to match the filtered faces to the names (e.g., Costen, Parker, & Craw, 1996; Fiorentini, Maffei, & Sandini, 1983), or observers were asked to match the filtered faces to nonfiltered versions of a small set of faces that was perpetually visible (e.g., Hayes et al., 1986). Second, the faces were typically shown for only brief durations (on the order of 100 msec) to avoid ceiling effects. Third, the studies used band-pass rather than low-pass filtered faces, as in the present experiments.

So it is not entirely clear how the results of these studies should be compared with the present results. These studies do, however, in conjunction with the present data, suggest at least a rough way of demonstrating the effect of distance on face identification, on the basis of fundamental image data. In particular, we calculated the contrast energy spectrum of each of our celebrities, filtered them with our  $F_1 = 31$  cycles/deg MTF, and then determined the total contrast energy remaining in the 11–20 cycles/face range at varying distances. Figure 15 shows the function relating this alleged *face identification energy* to distance. If the result is to be taken seriously, face identification should remain at 1.0 up to a distance of approximately 25 ft and then descend, reaching zero at a around 110 ft. This is close to what we observed in our Experiments 2–4. Although this correspondence occurs in part because we used the Experiment 2–4 MTF estimate to generate the function in Figure 15, the function is not based entirely on our data, and the correspondence may be useful as a foundation for further study of face identification energy.

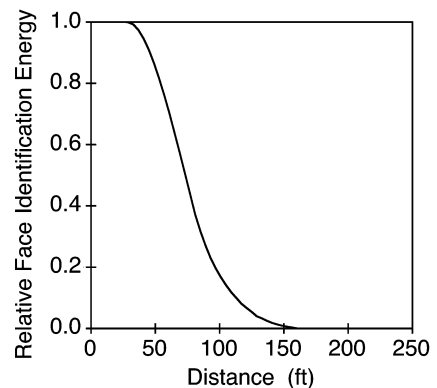
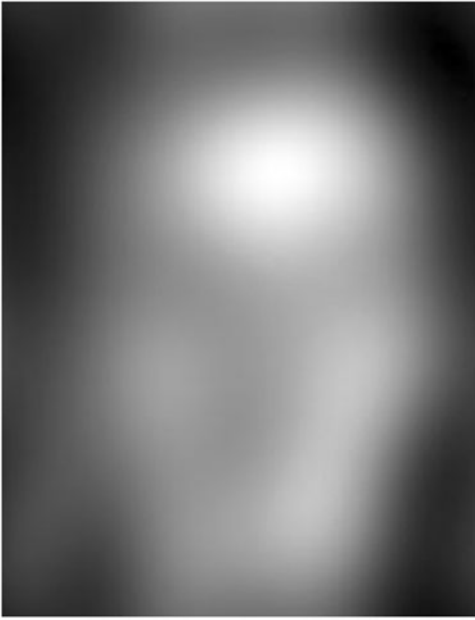


Figure 15. Total face identification energy (energy in 11–20 cycles/face region) as a function of face distance from the observer. Data are averaged over 64 celebrities.



**Figure 16.** Mystery celebrity filtered assuming a distance of 450 ft. The modulation transfer function has a cutoff parameter  $F_1$  of 42 cycles/deg.

**Memorizing unfamiliar faces.** We introduced this research by reference to a hideous crime in which an eye-witness, Arlo Olson, claimed to be able to recognize individuals whom he had previously watched committing the crime. Mr. Olson did not claim to know the culprits; thus, the identification procedure consisted of selecting suspects from a photographic lineup. This kind of recognition is different from the identification process studied here in Experiments 2–4, wherein observers attempted to identify people who were already familiar to them. Given that we have determined that the MTF for one task (e.g., matching, as in Experiment 1) is not necessarily the same as the MTF for another task (identification, as in Experiments 2–4), we cannot make an unqualified claim as to the degree of high spatial frequency information lost to Mr. Olson.

However, even taking a conservative view—that the MTF relevant to Mr. Olson’s recognition task is the less low-passed,  $F_1 = 42$  cycles/deg MTF from Experiment 1—the degree of degradation for an individual seen from a distance of 450 ft is considerable. Can you identify the well-known celebrity pictured in Figure 16? You can’t? If you can’t even identify him, it is unlikely that you would have been able to recognize him in a lineup had he been just a random, unfamiliar individual whom you had briefly watched commit a crime.

## EPILOGUE

On August 19, 1999, after an 11-day trial and 2 days of deliberations, a jury convicted the Fairbanks defendants of assault and murder. The defense attorney, the

expert witness, and even the prosecutor were astonished, because implicit in the jurors’ verdict was their belief that Arlo Olson must have accurately perceived and memorized the assailants from 450 ft away. So much for the persuasive effects of the Yankee Stadium analogy!

On February 24, 2003, the expert witness was sitting at his desk, writing a manuscript describing research that had been motivated by this case. Suddenly, out of the blue, an e-mail message appeared on his computer. It was from the defense attorney in the case! She wrote,

Greetings from Alaska!

We may have figured out why the jurors believed that Arlo Olson could identify [the defendants] at such a distance. A newspaper reporter turned journalism professor taught a class on our case last year. He had all his students interview witnesses, family members, jurors, etc. Four jurors have confirmed that they did their own experiment in broad daylight during a break. They had one or more jurors pace off a given distance and then they all stood and decided if they could recognize them! One lady said that because they could, it defeated everything that the defense attorneys said. Another said that he had bad eyesight and so even though he could not see, because the others could, he believed them. They all said it gave credibility to what Olson said. They were under the impression that they had permission to do this. The judge confirmed in an interview that he knew nothing about it.

That sure explains things! The article is coming out later this week. It was supposed to come out in the spring, but he moved it up so we could get the names of the jurors to file a motion for a new trial. I will send you a copy of the article when it is published.

Thought you might be interested.

## REFERENCES

- BAMBER, D. (1979). State trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*, **19**, 137-181.
- BERNSTEIN, D. M., ATANCE, C., LOFTUS, G. R., & MELTZOFF, A. N. (2004). We saw it all along: Visual hindsight bias in children and adults. *Psychological Science*, **15**, 264-267.
- BLAKEMORE, C., & CAMPBELL, F. W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, **203**, 237-260.
- BRACEWELL, R. N. (1986). *The Fourier transform and its applications* (2nd ed.). New York: McGraw-Hill.
- BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.
- BRUNER, J. S., & POTTER, M. C. (1964). Interference in visual search. *Science*, **144**, 424-425.
- CAMPBELL, F. W., & ROBSON, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, **197**, 551-566.
- COSTEN, N. P., PARKER, D. M., & CRAW, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics*, **58**, 602-612.
- DE VALOIS, R. L., & DE VALOIS, K. K. (1980). Spatial vision. *Annual Review of Psychology*, **31**, 309-341.
- DE VALOIS, R. L., & DE VALOIS, K. K. (1988). *Spatial vision*. New York: Oxford University Press.
- FARAH, M. J., WILSON, K. D., DRAIN, M., & TANAKA, J. N. (1998). What is “special” about face perception? *Psychological Review*, **105**, 482-498.
- FIORENTINI, A., MAFFEI, L., & SANDINI, G. (1983). The role of high spatial frequencies in face perception. *Perception*, **12**, 195-201.
- GEORGESON, M. A., & SULLIVAN, G. D. (1975). Contrast constancy: De-

- blurring in human vision by spatial frequency channels. *Journal of Physiology*, **252**, 627-656.
- GOSSELIN, F., & SCHYNS, P. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, **41**, 2261-2271.
- GRAHAM, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.
- HARLEY, E. M., BERNSTEIN, D. M., & LOFTUS, G. R. (2002, November). *Visual hindsight bias in object and face recognition*. Poster presented at the annual meeting of the Psychonomic Society, Kansas City, MO.
- HARLEY, E. M., CARLSEN, K., & LOFTUS, G. R. (2004). The "saw it all along" effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 960-968.
- HARLEY, E. M., DILLON, A., & LOFTUS, G. R. (2004). Why is it difficult to see in the fog? How stimulus contrast affects visual perception and visual memory. *Psychonomic Bulletin & Review*, **11**, 197-231.
- HARMON, L. D., & JULESZ, B. (1973). Masking in visual recognition: Effects of two dimensional filtered noise. *Science*, **180**, 1194-1197.
- HAYES, T., MORRONE, M. C., & BURR, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception*, **15**, 595-602.
- LOFTUS, G. R., & BUSEY, T. A. (1992). Multidimensional models and iconic decay. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 356-361.
- LOFTUS, G. R., JOHNSON, C. A., & SHIMAMURA, A. P. (1985). How much is an icon worth? *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 1-13.
- LOFTUS, G. R., & RUTHRUFF, E. R. (1994). A theory of visual information acquisition and visual memory with special application to intensity-duration trade-offs. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 33-49.
- MORRISON, D. J., & SCHYNS, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, **8**, 454-469.
- OLIVA, A., & SCHYNS, P. G. (1997). Coarse blobs or fine edges: Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, **34**, 72-107.
- OLZAK, L. A., & THOMAS, J. P. (1986). Seeing spatial patterns. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 7-1 to 7-56). New York: Wiley.
- O'TOOLE, A. J., WENGER, M. J., & TOWNSEND, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (pp. 1-38). Mahwah, NJ: Erlbaum.
- PALMER, J. C. (1986a). Mechanisms of displacement discrimination with and without perceived movement. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 411-421.
- PALMER, J. C. (1986b). Mechanisms of displacement discrimination with a visual reference. *Vision Research*, **26**, 1939-1947.
- PARISH, D. H., & SPERLING, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, **31**, 1399-1415.
- PELLI, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.
- REINITZ, M. T., & ALEXANDER, R. (1996). Mechanisms of facilitation in primed perceptual identification. *Memory & Cognition*, **24**, 129-135.
- ROBSON, J. G. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America*, **56**, 1141-1142.
- SCHYNS, P. G., & OLIVA, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentation. *Cognition*, **69**, 243-265.
- SHLAER, S. (1937). The relation between visual acuity and illumination. *Journal of General Psychology*, **21**, 165-188.
- TOWNSEND, J. T., SOLOMAN, B., & SMITH, J. S. (2001). The perfect gestalt: Infinite dimensional Riemannian face spaces and other aspects of face perception. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (pp. 39-82). Mahwah, NJ: Erlbaum.
- TULVING, E., & SCHACTER, D. L. (1990). Priming and human memory systems. *Science*, **247**, 301-306.
- VAN NES, R. L., & BOUMAN, M. A. (1967). Spatial modulation transfer in the human eye. *Journal of the Optical Society of America*, **57**, 401-406.

#### NOTES

1. In this article, we define face size as face *height*, because our stimuli were less variable in this dimension than in face *width*. Other studies have defined face size as face *width*. A face's height is greater than its width by a factor of approximately 4/3. When we report results for other studies, we will transform relevant numbers, where necessary, to reflect face size defined as height.
2. It can easily be shown that the best-fitting zero-intercept slope of  $y$  against  $x$  in an  $x$ - $y$  plot is estimated as  $\Sigma xy / \Sigma x^2$ . In this instance, the estimated zero-intercept slopes were almost identical to the unconstrained slopes.
3. Throughout this manuscript,  $x \pm y$  refers to a mean,  $x$ , plus or minus a 95% confidence interval.
4. On the basis of data from Experiments 2 and 3 and a celebrity familiarity survey completed by lab members, less familiar celebrities from the prior experiments were replaced with new, more familiar celebrities.

(Manuscript received July 24, 2003;  
revision accepted for publication March 31, 2004.)