Cognitive Science 28 (2004) 259–287

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Object recognition and Random Image Structure Evolution

Javid Sadr*, Pawan Sinha

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology,
45 Carleton Street, Room E25-201, Cambridge, MA 02142, USA

Received 19 September 2002; received in revised form 4 September 2003; accepted 25 September 2003

Abstract

We present a technique called Random Image Structure Evolution (RISE) for use in experimental investigations of high-level visual perception. Potential applications of RISE include the quantitative measurement of perceptual hysteresis and priming, the study of the neural substrates of object perception, and the assessment and detection of subtle forms of agnosia. In simple terms, RISE involves the measurement of perceptual and/or neural responses as visual stimuli are systematically transformed—in particular, as recognizable objects evolve from, then dissolve into, randomness. Points along the sequences corresponding to the onset and offset of subjects' percepts serve as markers for quantitatively and objectively characterizing several perceptual phenomena. Notably, these image sequences are created in a manner that strictly controls a number of important low-level image properties, such as luminance and frequency spectra, thus reducing confounds in the analysis of high-level visual processes. Here we describe the RISE paradigm, report the results of a few basic RISE experiments, and discuss a number of experimental and clinical applications of this approach.

Keywords: Perception; Vision; Object recognition; Face recognition; Detection; FMRI; MEG; Image processing; Psychophysics, Methods; Neuroimaging; Priming; Pattern recognition; Clinical; Hysteresis

1. Introduction

Conventional studies of high-level visual perception, and especially those of object recognition, typically characterize behavioral and/or neural responses to stimulus images depicting various objects or scenes. In addition to numerous behavioral experiments, examples of such

* Corresponding author. Present address: Vision Sciences Laboratory, Department of Psychology, Harvard University, 33 Kirkland Street, 7th Floor, Cambridge, MA 02138, USA. Tel.: +1-617-495-3884; fax: +1-617-495-3764.
E-mail addresses: sadr@wjh.harvard.edu (J. Sadr), sinha@ai.mit.edu (P. Sinha).

0364-0213/$ – see front matter © 2003 Cognitive Science Society, Inc. All rights reserved.
doi:10.1016/j.cogsci.2003.09.003

studies include electrophysiological studies of inferotemporal (IT) cortical neurons which assess cells' responses to images of objects such as faces, hands, toilet-brushes, and so on (Desimone, Albright, Gross, & Bruce, 1984; Perrett, Hietanen, Oram, & Benson, 1992), and functional neuro-imaging studies that attempt to identify object-specific cortical areas (e.g., Kohler, Kapur, Moscovitch, Winocur, & Houle, 1995; Puce, Allison, Gore, & McCarthy, 1995). However, one can argue that functional interpretations of the responses observed in these studies are based on what amounts to a sparse and, more importantly, relatively unsystematic sampling of the space of all possible stimulus images. When the dependent variable is sampled in this manner, interpreting and extrapolating from the obtained data amount to ill-posed problems. (One can imagine the difficulty of plotting a curve as a function of an independent variable x if too few data points exist, if these points are distributed too unevenly along the x axis, or, worse, if there is no formal metric by which to represent the variations in x along the abscissa.) Further, in order to interpret the results of such experiments, it is of great importance to ensure that the seemingly high-level effects of interest are not contaminated by low-level confounds (e.g., covariation in stimulus contrast, power spectrum, etc.). We have developed a flexible new approach, called Random Image Structure Evolution (RISE), that addresses these issues by probing responses to dense sets of visual stimuli that are sampled in a systematic fashion and controlled for a number of important low-level properties (in particular, spatial frequencies, luminance, and contrast).

As depicted schematically in Fig. 1, images can be thought of as points lying in a high-dimensional space, where each dimension corresponds to one of the ways in which images may vary. (For instance, a 100 × 100 pixel grayscale image may be represented as a point in a 10,000 dimensional space, with each dimension corresponding to the luminance of a pixel.) RISE enables the generation and experimental presentation of sets of visual stimuli sampled from this space at various distances from any image of interest. These sets of images can be thought of as trajectories through image space passing through pre-selected or even random images. By enforcing "continuity" in its sample set, and by providing a simple metric that relates the stimulus images to one another, RISE allows for a meaningful comparison of responses across the entire set of images. Thus, one can examine how responses change in moving from one point along the trajectory to its neighbor and correlate this change with the incremental image-level change. Discontinuities or pronounced nonlinearities (i.e., "categorical" changes) in response while moving along a continuous trajectory may be of particular significance since these may signal the onset of high-level visual and cognitive events. As explained above, an unsystematic selection of points from the image space does not allow such analysis, in so far as one would not be able to represent such changing response profiles with respect to a quantitative, ordered series of values of a (continuous) independent variable.

Working with systematically sampled and continuous trajectories rather than isolated points can greatly enhance one's experimental and analytical repertoire. Changes in a variety of attributes may be measured as a function of the position along these image trajectories; in addition to basic perceptual responses (e.g., the onset of object recognition), these attributes may include measures of neural activity as well as theoretical indices of the information content of the stimuli. By analyzing the mutual correlations of these behavioral, neural, and image-based variables while simultaneously removing as many confounding factors as possible, one may obtain information critical for answering a host of important questions in high-level vision.

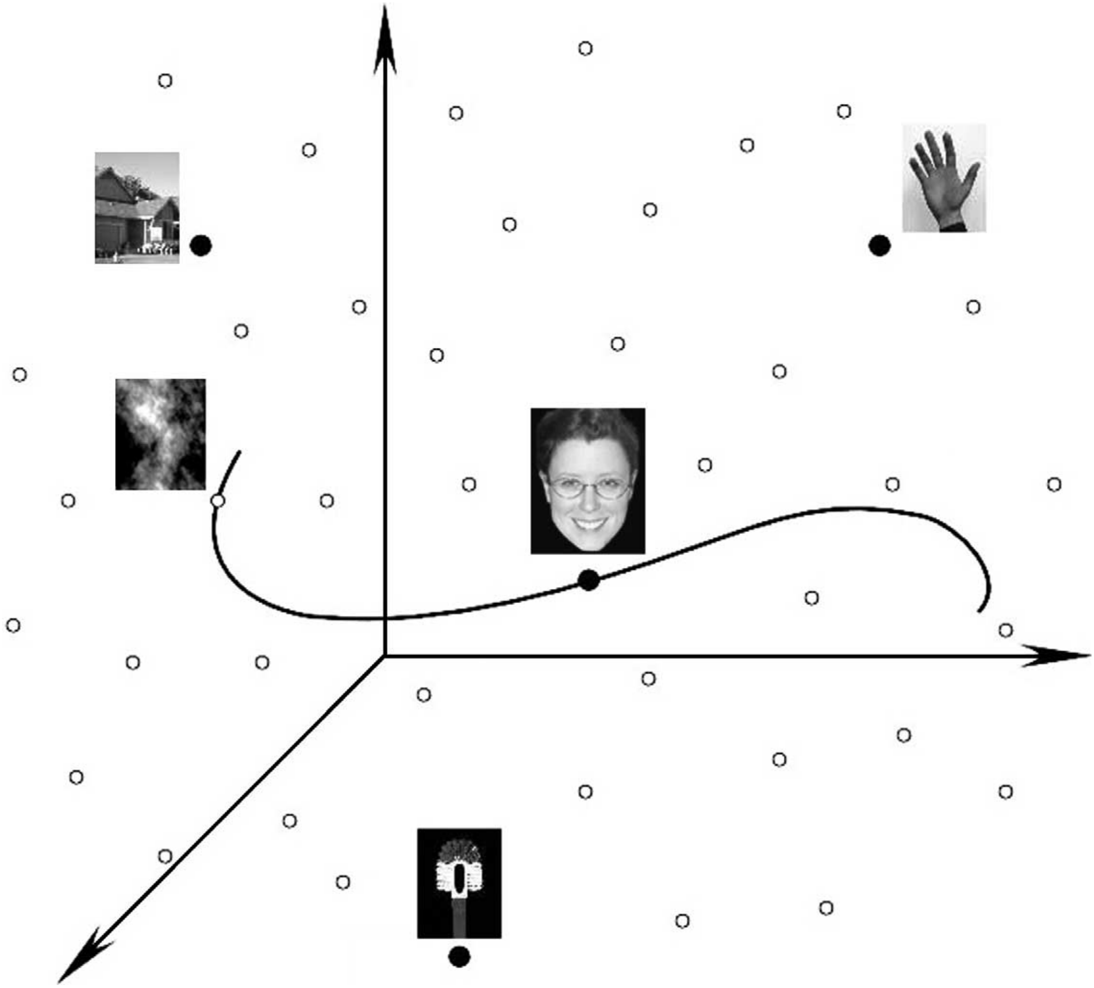

Fig. 1. Conceptually, a given image can be said to correspond to a point lying in a high-dimensional "image space," as shown here schematically. Specific images of objects and scenes, of the kind used in most recognition studies constitute a very sparse subset of points in this space (filled circles). Making inferences about high-level visual processes based on responses measured at a few, possibly quite disparate points is something of an underdetermined problem. The motivation behind RISE is to overcome this problem by studying perceptual and neural processes along continuous trajectories (black curve) passing through specific points of interest in this image space. Such trajectories, then, are simply image sequences depicting well-controlled visual transformations.

This is the key motivation underlying the RISE paradigm. Here, we describe the basic methodology of RISE and, as a starting point, the results of RISE experiments concerned with such phenomena as perceptual hysteresis and priming. We also discuss a number of other interesting and important ways in which the RISE paradigm might contribute to the study of high-level vision, including the characterization of the neural substrates of object perception, the quantitative assessment of priming, and the exploration of perceptual learning and development, not to mention the study of top–down influences on early visual areas.

2. The RISE paradigm

2.1. Stimulus image processing

RISE can be thought of as a specific type of morphing (Benson & Perrett, 1993; Busey, 1998; Shelton, 1998) or image degradation procedure (e.g., Dolan et al., 1997; James, Humphrey, Gati, Menon, & Goodale, 2000; Snodgrass & Corwin, 1988; see also Harmon & Julesz, 1973). A very simple version of this technique proceeds by exchanging randomly selected pairs of pixels in an image. As these replacements accumulate, the original image dissolves into a random field. This procedure can be carried out in reverse order as well, allowing the image transformation to be displayed backward as well as forward. As such, the first half of a sequence ("onset" subsequence) could show the image emerging from a random field, while the second half ("offset" subsequence) shows the image disappearing back into randomness. (See Fig. 2 for a sample sequence.) Thus, the two extremes of a complete sequence are random patterns while the midpoint is a fully constituted image; this may be thought of as a continuous trajectory in image space passing through the original image as its midpoint. The sizes of the flipped regions and the spatial extents of the transpositions (with small extents leading to local structure randomization) are under the experimenter's control. Of course, the original image may depict anything of interest (e.g., an object, face, scene, abstract shape, etc.).

Besides being computationally simple to implement, this procedure possesses a very attractive characteristic: it precisely maintains global photometric attributes such as luminance and color histograms. This avoids confounding of the experimental results with changes in these low-level attributes. This technique has one shortcoming, however, which it shares with other approaches in which images are scrambled or partially occluded (e.g., Grill-Spector et al., 1998; James et al., 2000) or are subjected to additive or multiplicative noise (e.g., Rainer & Miller, 2000): it does not preserve the frequency spectrum of the source image. That is, an unintended side-effect of these techniques is the disruption of the original distribution of spatial frequencies (ranging from sharp edges to smooth gradients, each with a specified orientation) that compose the image of interest.

In image scrambling, the increasingly fragmented or pixelated images are constituted progressively more by higher frequencies, especially those coinciding with the cardinal axes—that is, the more scrambled images are composed by sharper edges, particularly ones oriented vertically and horizontally. One can readily imagine the many sharp edges that are produced in an image when one divides a picture of a relatively smooth object (e.g., a face) into small squares and then randomly rearranges these squares. Unfortunately, not only is one very unlikely to conserve the original frequency spectrum by blurring the scrambled images or by adding higher frequencies to the original image (Grill-Spector et al., 1998), these may not be particularly desirable manipulations in the first place.

In the case of noisy images, even if these are created by linear interpolation between a source image and a randomized image with an identical frequency spectrum (Rainer & Miller, 2000), there can be no expectation of conserving the frequency spectrum at anywhere but the extremes of the transformation. To see this, imagine a linear interpolation (i.e., a succession of weighted averages) between two complementary black-and-white checkerboards, identical in terms of luminance, contrast, and frequency spectrum. The result will be a series of grayscale

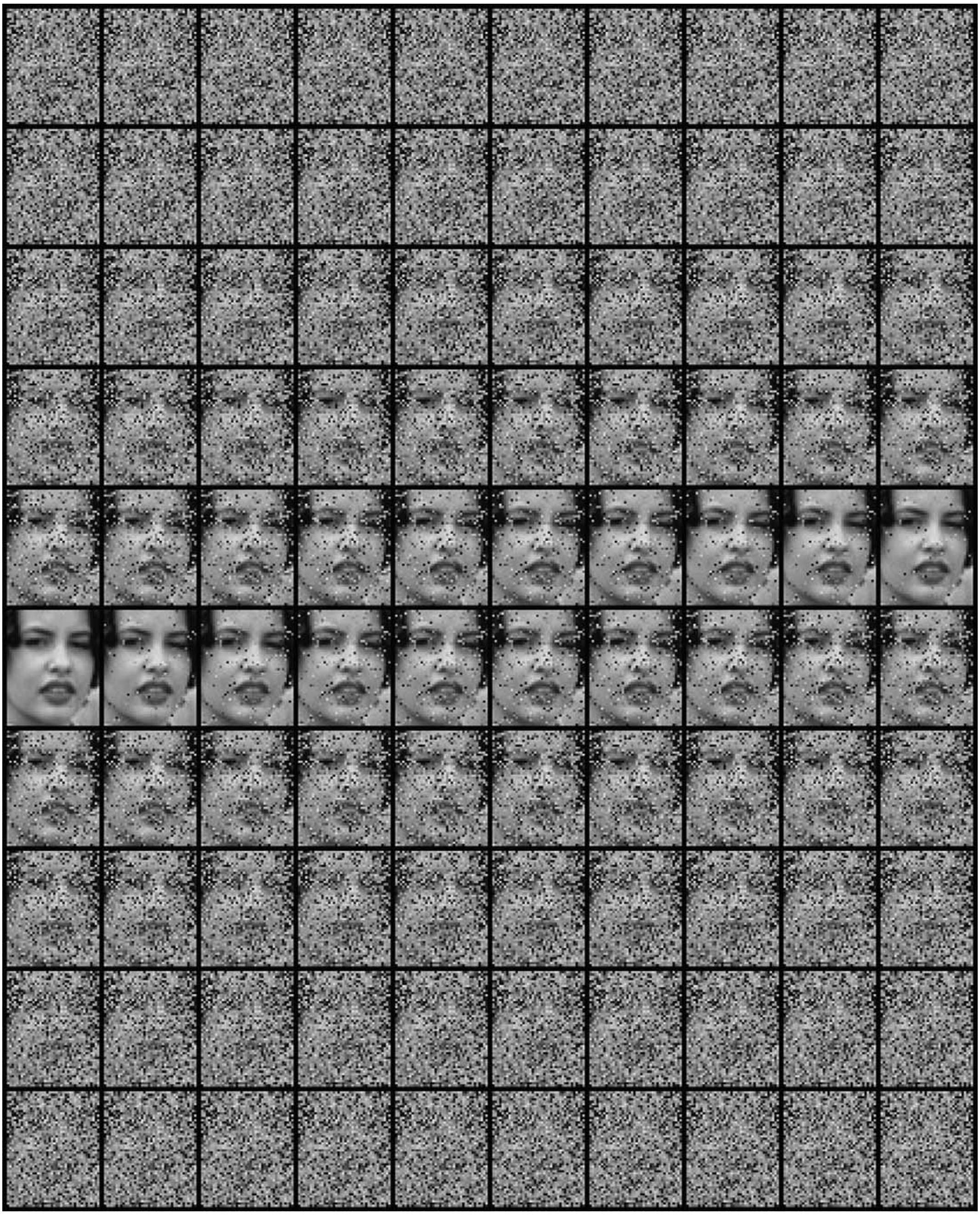

Fig. 2. A sample RISE sequence generated by pairwise exchanges of image regions (here single pixels). A simple presentation of these images would proceed in raster order (i.e., from left to right and top to bottom). The source image appears in the first column of the sixth row.

images, and, most strikingly, the mid-point image will be a uniform gray field. Clearly, this transformation does not preserve the frequency spectrum of the original images. Perhaps even more importantly, the luminance contrast of the images neither remains constant nor varies monotonically: along the sequences, there is first a decline in contrast from the starting image to the mid-point, then an increase back up to the contrast of the final image. (This issue is discussed further below.)

There are a number of reasons why one would be interested in controlling (or, conversely, selectively and actively manipulating) the frequency spectrum of the stimuli in experiments interested in higher level vision. To begin with, the particular structure of the frequency spectrum of object images, and/or the relationship of these spectra to the complex manner in which the human visual system filters and processes said images, can have very pronounced effects on high-level perception. For example, it has been shown that, for the successful recognition of faces, a relatively small range of frequencies from approximately 8 to 16 cycles across the face are of surprisingly great importance (e.g., Costen, Parker, & Craw, 1996; Nasanen, 1999; Schyns, Bonnar, & Gosselin, 2002; and others). That is, there is a significant decline in subjects' ability to recognize a face if in the stimulus image the information in this frequency band is disrupted; conversely, recognition performance remains quite good when this frequency band is preserved while other spatial frequencies are disrupted. (A detailed review of this topic, looking also at objects and scenes and at questions of visual perception, attention, and representation (e.g., the use of fixed vs. flexible coarse-to-fine strategies of spatial scale usage) can be found in Morrison & Schyns, 2001.) Moreover, it has been suggested in this case that this effect is likely a result of the properties of the human visual system rather than an intrinsic characteristic of the images themselves (Gold, Bennett, & Sekuler, 1999), so that an analysis of the stimulus images is unlikely to indicate *a priori* which frequency bands are more or less important to conserve for recognition. In much the same vein, there has been work studying the role of so-called spatial frequency channels for reading and letter identification (e.g., Gold et al., 1999; Legge, Pelli, Rubin, & Schleske, 1985; Majaj, Pelli, Kurshan, & Palomares, 2002), and this too has explored the influence of spatial frequencies through both psychophysics and more bottom–up image-based analysis.

Further, it has been of considerable interest to explore the relationship between the spatial (as well as temporal) frequencies of visually presented stimuli and the corresponding neural activity elicited within various structures in the visual system (e.g., Singh, Smith, & Greenlee, 2000). Such work reinforces the notion that the spatial frequency content of images interacts with the architecture of the visual system in such a way as to produce complex patterns of neural activity observed throughout the brain even during relatively simple visual processing. There has also been work discussing the relationship between spatial frequencies and high-level visual processes such as basic-level object categorization, for example through the study and modeling of the role of infants' relatively weak visual acuity (which essentially filters out higher frequencies making images appear more blurry) in biasing categorical perception (French, Mermillod, Quinn, Chauvin, & Mareschal, 2002).

Given these considerations, a preferred alternative approach to the very simple image scrambling implementation of RISE is to first perform an analysis of the frequency spectrum (i.e., a Fourier analysis) of the source image, then manipulate the spatial structure of the image without altering its original power spectrum (as well as the overall luminance and contrast of the image). In the Fourier domain, this can be done by altering what is called the phase spectrum while retaining the amplitude (or power) spectrum. In fact, it has been shown that much of the information specifying natural image structure lies in the global phase spectrum (Oppenheim & Lim, 1981), so that randomizing the phase spectrum has the effect of degrading the spatial structure of an image. (Further, replacing the phase spectrum of an image with that of another image results in an image that resembles the donor of the phase rather than of the amplitude spectrum; however, the lower-level attributes of the resulting image may better resemble the donor of the amplitude spectrum.) As such, an alternative to the scrambling technique described above is to manipulate the source image in the Fourier domain, progressively transforming the phase while holding constant the amplitude spectrum (Sadr & Sinha, 2001a, 2001b). In the case of the onset portion of a RISE sequence, then, the perfect image evolves from a random-seeming starting image that has been constructed using a random phase spectrum combined with the amplitude spectrum of the original image. The onset subsequence is achieved through progressive transformation of the random phase spectrum into that of the perfect image, and the offset subsequence is simply this process in reverse. (See Fig. 3 for a sample sequence.) For those interested in perhaps implementing this technique or a variation thereof, a more detailed description follows below. First, however, a brief overview of relevant concepts in Fourier analysis may be in order.

As a starting point, the Fourier transform can be thought of as the decomposition of a signal, e.g., a sound or an image, into a set of simple constituent signals (Bracewell, 2000; De Valois & De Valois, 1988). One can imagine, for instance, how a sound wave might be composed of a number of individual sine waves, each with its own frequency (e.g., 1 Hz, 2 Hz, etc.), amplitude, and relative offset (angle or phase), along with a non-periodic (0 Hz or “DC”) baseline value. Fig. 4 shows how such a decomposition might look for a 2-D signal such as a 100 × 100 pixel image (Fig. 4a). Here, the Fourier transform, a collection of complex (i.e., real and imaginary) coefficients, is further decomposed into its corresponding real amplitude and phase values (Fig. 4b and c, respectively).

As depicted, the origin of the amplitude and phase spectra is situated at the center, with the horizontal and vertical axes corresponding to horizontal and vertical spatial frequencies, respectively. Here, for example, the vertical axis spans vertical spatial frequencies from -49 to 50 Hz, where Hz denotes cycles across the width or height of the image. (Notice that the highest frequency coincides with one half of the resolution of the image. In one sense, the finest edge that may reasonably be assessed by the Fourier transform is that between a pair of pixels, and the limit here is 50 pairs, or 50 cycles along the image.) For real-valued signals, “negative” frequencies simply represent complex conjugates of the positive frequencies, so both the amplitude and phase spectra exhibit an inherent symmetry.

With amplitude and phase spectra in hand, one may reproduce the original signal, or a close approximation thereof, using the inverse Fourier transform. Notably, if one manipulates the amplitude and/or phase spectra prior to performing the inverse transform, these changes will be reflected in the reconstructed signal. A common example is the creation of blurry images by the attenuation of high frequency values in the amplitude spectrum; boosting these values produces a sharper (and often noisier) image. Here, the focus is on manipulating the phase spectrum, and we present a simple illustration to provide a sense of how such changes in the Fourier domain may be reflected in the image domain. Fig. 4d–f depict the output of the inverse Fourier

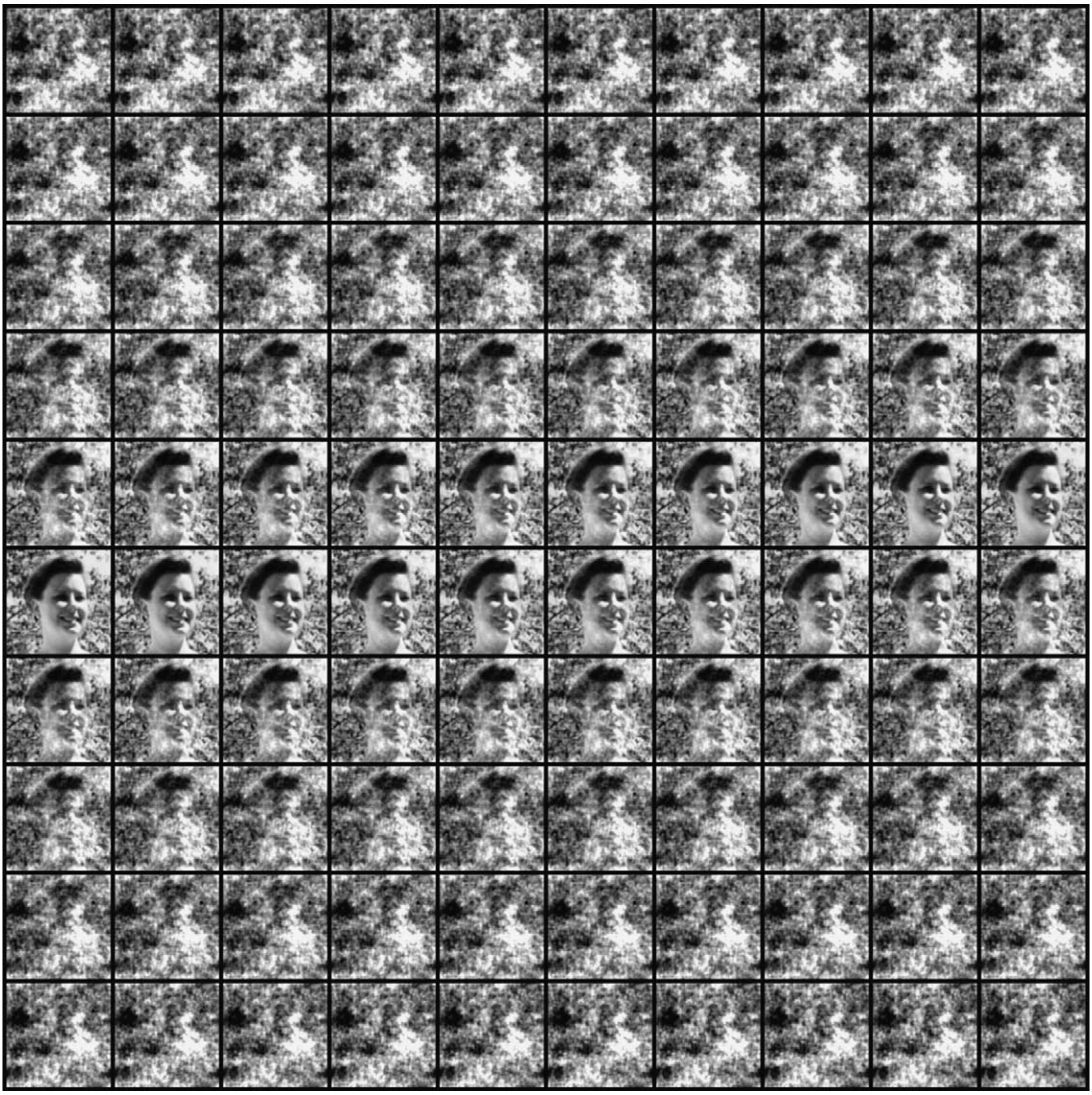

Fig. 3. A sample RISE sequence generated by progressive degradation of the phase matrix of the source image. Although the technique progressively disrupts the spatial structure of the image, important low-level image properties of the original image, such as the spatial frequency spectrum and overall luminance, are perfectly preserved.

transform following the randomization of phase values in specific frequency bands (1–16 Hz, 17–32 Hz, and 33–48 Hz, respectively). Whereas the output of the inverse Fourier transform using the original phase and amplitude spectra would give an image virtually identical to the original, each of these phase-manipulations results in a distinct disruption of the structure of the image specific to the targeted frequency bands.

To perform the image transformations in RISE, then, we begin by performing a Fourier transform, typically a discrete Fast Fourier Transform (FFT), of the original image of interest.

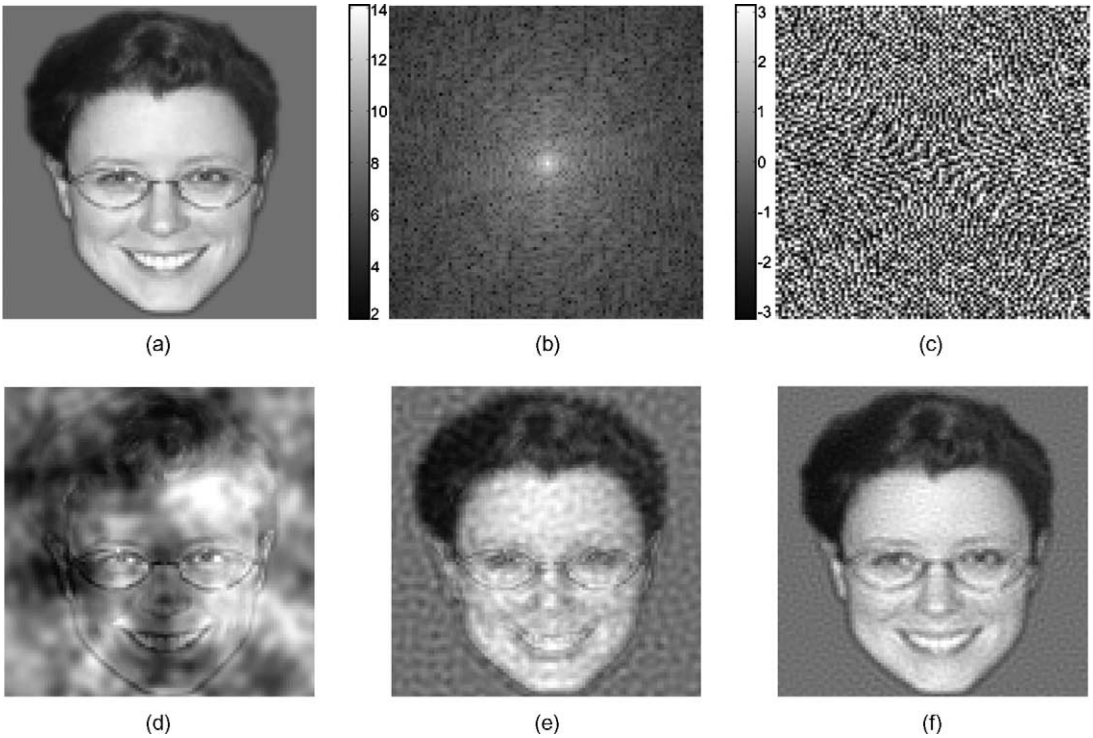

Fig. 4. Simple illustration of Fourier analysis and phase randomization. Top row: (a) original image, (b) log of amplitude spectrum, and (c) phase spectrum (values range from $-\pi$ to π radians). Bottom row: output images from inverse Fourier transform given original amplitude spectrum and modified phase spectrum randomized in the (d) 1–16 Hz, (e) 17–32 Hz, and (f) 33–48 Hz frequency bands.

We extract the amplitude and phase spectra from the complex elements of the FFT of the image. Each of the images in the RISE sequence is produced by way of an inverse FFT (IFFT) performed on the combination of the original amplitude spectrum with a modified phase spectrum. As such, in our phase-manipulation implementation of RISE, all images in the sequence have identical amplitude spectra and luminance.

In the simplest case, the original phase spectrum is gradually transformed into a random one consisting of a similar distribution of phase values. Although a simple linear interpolation may be used to achieve this transformation, we first perform a randomized operation on the target (random) phase spectrum that, for approximately half of the elements, circumvents the zero-crossings that would otherwise occur (i.e., the preponderance of zero and near-zero phase values at and around the 50% interpolation level, resulting in a non-uniform phase distribution with a marked mode at zero). This feature of the technique, discussed again further below, serves to maintain the distribution of phase values over the course of the transformation and, ultimately, to help control the contrast of the images produced. First, we randomly select half of the target phase values. Second, for each of these values, we either add or subtract 2π. The choice between adding or subtracting 2π is made in such a way as to yield a target value that is of the same sign as the corresponding element in the original phase spectrum, thus no longer requiring these phase elements to pass through a value of zero during the interpolation between their original and target values. For example, if a certain element of the original phase spectrum has a value of 0.5 and is being interpolated to a final value of -0.5 in the randomized spectrum, we essentially flip a coin and decide whether or not to add 2π to the final value (0.5) in order to circumvent the zero-crossing (i.e., by interpolating from 0.5 to $-0.5 + 2\pi$ rather than to -0.5). For interpolations in the other direction (e.g., from -0.5 to 0.5), we would randomly choose whether or not to subtract 2π from the target phase value.

Finally, there are two additional considerations regarding the target phase spectrum. First, the phase value corresponding to the zero frequency (or DC component, seen at the center of the spectra shown in Fig. 4b and c) is not changed from that of the original. Second, because all real-valued signals (e.g., normal sounds and images) have Fourier spectra that are symmetric, we ensure at all times that the modified phase spectra retain this symmetry. In this way, the IFFT does not produce signals that contain imaginary values and, consequently, images that have undergone inadvertent changes in power spectrum.

When creating a number of RISE sequences for use in an experiment, it may in some cases be desirable to perform certain pre-processing procedures on the source images. For example, one may wish to ensure that all of the object images in an experiment (or experimental block) have equivalent frequency spectra so that both within and across RISE sequences only the phase spectra vary. One way of doing this is to collect the amplitude spectra of all the original object images and from them create an average amplitude spectrum (e.g., Bracewell, 2000); a new set of source images could then be constructed by combining the average amplitude spectrum with each of the unmodified phase spectra and performing the IFFT. It is worth noting that the IFFT may give results beyond the range of acceptable luminance values (i.e., luminance values less than zero or greater than those supported by the display); a simple linear shift and rescaling of the luminance values can be performed to bring them into the correct range, but precisely the same operation should be performed over the entire set of normalized source images. In a similar fashion, one could precede the RISE image processing by normalizing the luminance histograms of all the source images, a potentially important pre-processing step if the original images vary a great deal in their luminance and contrast. Typically, this would be done before rather than after the normalization of the amplitude spectrum. Finally, it may be of interest to note that one can choose to have all of these RISE transformations converge to a single final image. This is done quite simply, by choosing one random phase spectrum to serve as the common end-point for all the interpolations.

For experimental purposes, we ensure that the random phase spectrum chosen for the image sequence generation, in combination with the random addition and subtraction of 2π in the interpolation, ultimately results in the monotonic evolution and degradation of the image (e.g., monotonic decrease and increase in absolute (L1) or the sum of squared differences (L2 distance) from the source image) during the onset and offset subsequences, respectively (Fig. 5). More importantly, however, while the technique described above may allow a small drift in image contrast, we nevertheless ensure that any such drift is monotonic across the desired sequence of images chosen for presentation in a RISE experiment. (This can be done by selecting a different final random phase spectrum, for example, and/or by reassigning the random additions and subtractions of 2π.) By comparison, another image degradation technique based on simple interpolation of phase spectra (Rainer, Augath, Trinath, & Logothetis, 2001) has

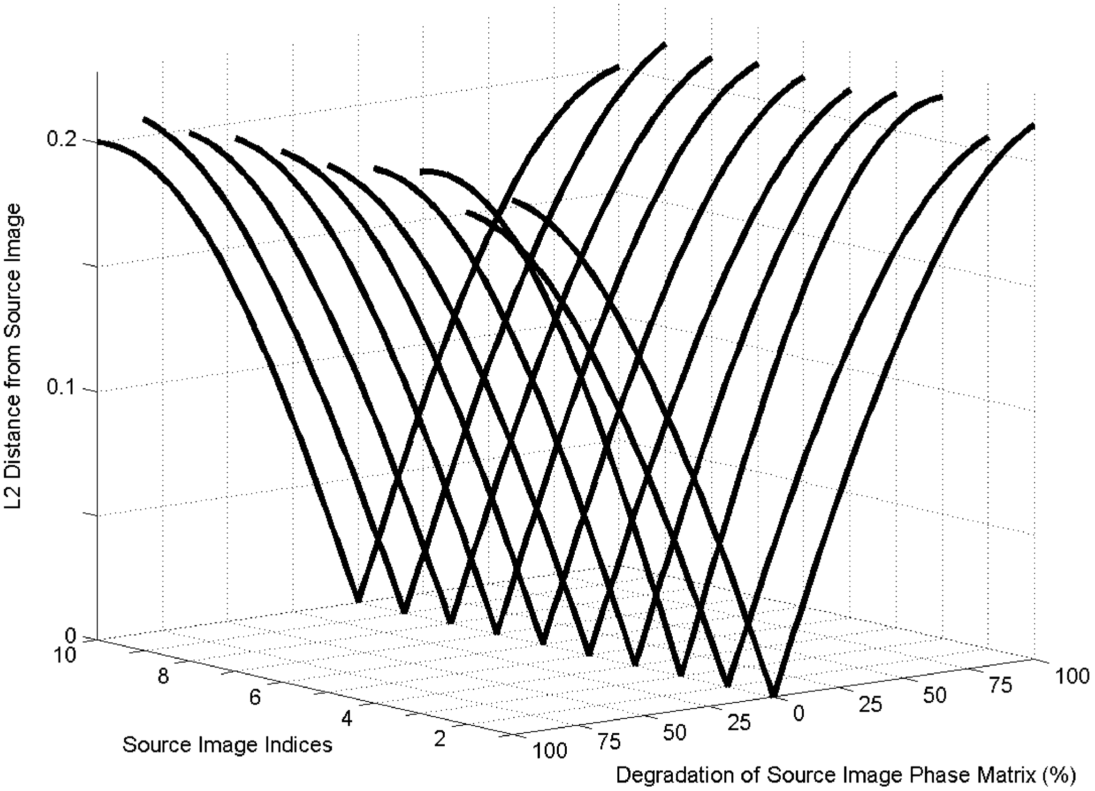

Fig. 5. Ten RISE sequences were generated, each based on 1 of 10 source images common objects. Here are plotted the L2 distances (i.e., sum of squared differences) between each of the original source images and the images in its associated RISE sequence. It can be seen that RISE image processing can be performed in such a way as to ensure monotonic evolution and degradation within each of the onset and offset subsequences, respectively. Notice that, because the source images appear in non-degraded form at the midpoint of the RISE sequences (i.e., at the transition between the onset and offset subsequences), here the L2 distance is appropriately 0.

received strong criticism (Dakin, Hess, Ledgeway, & Achtman, 2002) due to the fact that it produces image sequences with marked and non-monotonic changes in contrast. If one does not compensate for changes in the distribution of the phase values, and specifically an increase in the number of near-zero phase values toward the middle of the interpolation sequence, the resulting images develop bright corners and lose contrast elsewhere. This results in an important low-level confound in the interpretation of data collected using such image sequences.

Note that in addition to depicting the evolution/degradation of an image from/to a random counterpart, with trivial variation this technique may also be used to "morph" between two or more source images. In terms of the high-dimensional image space discussed above, such transformations would correspond to trajectories connecting two or more pre-selected points of interest. As above, the transformation of the phase matrices may be brought about by various methods, such as by the modified interpolation scheme or by random, accumulating substitution of elements. If the source images are first normalized in terms of their global power spectra and luminance, these will be held constant throughout the morph sequence, with the only change being in the underlying phase.

2.2. Basic experimental paradigm

In the simplest case, using RISE sequences in which one pre-selected object image emerges from and then dissolves back into a random field, one can obtain quantitative measures of at least two important aspects of an observer's percepts. The first of these may be called the perceptual onset point, the position along the initial half of the RISE sequence where the observer is first able to (correctly) identify the emerging image. The second measurement is the perceptual offset point, the position along the second half of the RISE sequence beyond which the observer is no longer able to recognize the target image. As discussed below, these two measurements can be of great use in the study of numerous aspects of high-level vision, but before proceeding to explore these potential applications, we first discuss how RISE sequences may actually be presented to observers in an experimental setting.

Passive viewing of RISE sequences may be adequate in certain experimental settings. For example, one can think of RISE image sequences as playing much the same role as the simpler time-varying visual stimuli passively viewed in electrophysiology experiments designed for subsequent reverse correlation analysis (e.g., DeAngelis, Ohzawa, & Freeman, 1993). However, for most behavioral and functional neuro-imaging experiments overt responses will be required to assess subjects' perceptual experiences. In fact, in such cases it is also important to establish objective verification of the subjects' perceptual reports. If subjects are naive as to the image that will appear in a RISE sequence and are instructed to identify the image as soon as possible during the onset subsequence, the subjects' correct identification of the object presented can serve to validate subjects' verbal reports of perceptual onset. Objectively verifying perceptual offset, however, requires a slightly more sophisticated approach.

A solution we propose is to insert distractor images throughout the original RISE offset subsequence (Fig. 6). These distractors are taken from RISE sequences created from other source images, with each distractor image chosen to be at a level of degradation between those of the preceding and following images. As such, the resulting "mixed" RISE offset subsequence depicts a series of images of progressively greater degradation, but each of these may or may not be degraded versions of the original "target" image. As this mixed offset subsequence proceeds, there will come a point when subjects can no longer recognize and reliably identify the presence of the target image from among the distractors; this point serves as the measure of perceptual offset. Obviously, with distractors present, subjects must actively report their frame-by-frame percepts and cannot continue merely to automatically name the target item at each frame of the offset subsequence. Moreover, subjects also can not rely on emerging noise patterns as a means of identifying each item, because all items evolve toward the same degraded image.

What follows is a simple illustrative experiment using the RISE image processing and stimulus presentation techniques described above. Here, we use RISE to measure the perceptual onset and offset points of subjects for a small set of object images, in a manner resembling the ascending and descending series in the method of limits. In the process, we obtain objective and quantitative measures of perceptual hysteresis that are not confounded either with low-level properties of the images or with the processes of image presentation or response collection.

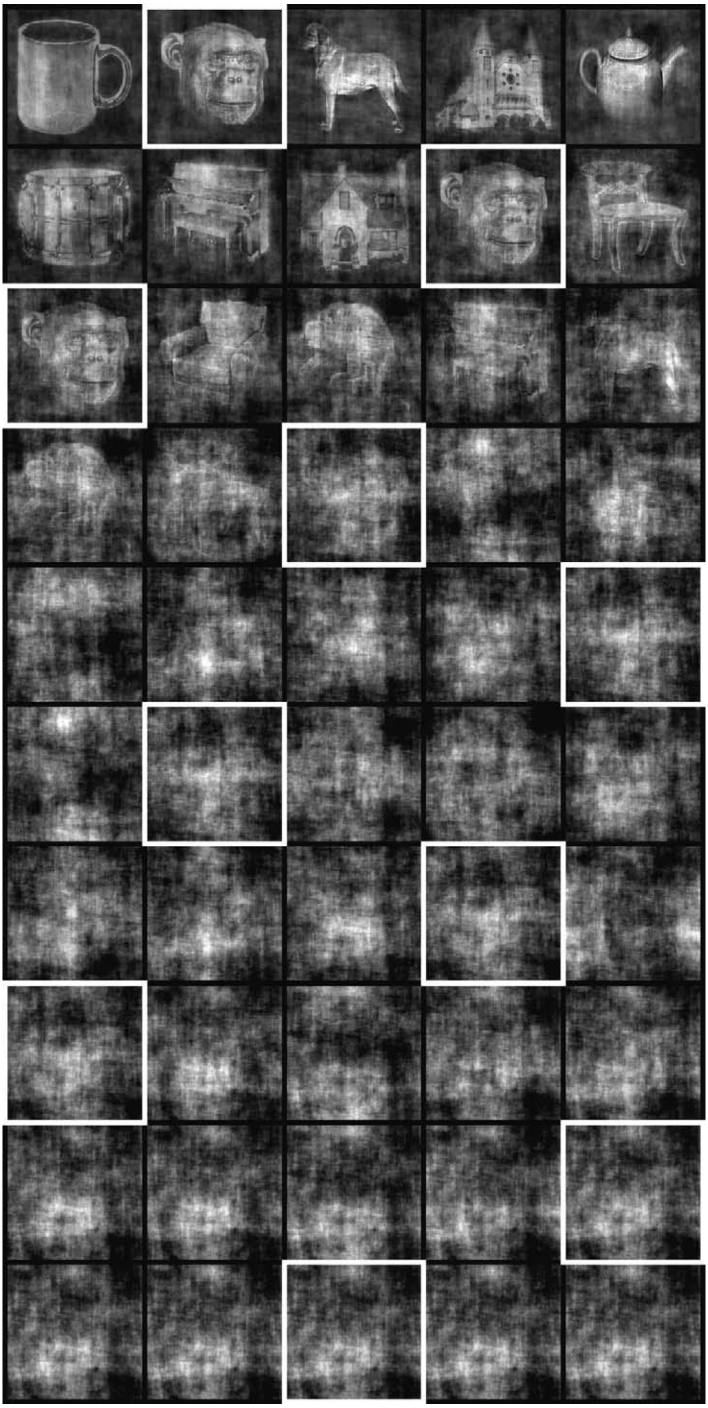

Fig. 6. A sample “mixed” RISE offset subsequence incorporating distractors. Here we have indicated the target object by framing it in white. Throughout the sequence, the images have precisely the same luminance and frequency amplitude spectrum (Fourier magnitude). Here, they also gradually converge to a common random phase spectrum.

2.3. *Experiment A: perceptual onset, offset, and hysteresis*

2.3.1. *Aim*

The purpose of this experiment is to illustrate the use of RISE in objectively assessing onset and offset points along a number of image trajectories and in obtaining objective and quantitative measures of the effects of perceptual hysteresis.

2.3.2. *Methods*

2.3.2.1. Participants. Four students from the Massachusetts Institute of Technology (MIT) participated in this study. All subjects provided informed consent and received payment according to MIT guidelines.

2.3.2.2. Stimuli. Five RISE sequences, of the "mixed" distractor variety, were generated using the phase-manipulation technique described above. Each of the sequences depicted the randomly interleaved evolution of one "target" object and 10 distractors. These sequences were generated from 255 × 255 pixel grayscale images of easily identified objects (see Fig. 7 for the five target objects), and the luminance histograms and amplitude spectra of all the source images were first normalized, as described above, prior to the creation of the RISE sequences. Each onset and offset subsequence consisted of 75 images ranging, in steps of 2.5%, from 50 to 85% interpolation of the source and random phase spectra, where 100% interpolation corresponds to the unaltered (though pre-processed) source image. Starting from the first frame of these mixed sequences, each group of five frames consisted, in random order, of one target-object- and four distractor-object-based RISE frames, and, in a slight departure from the distractor technique described above, each of the 15 groups of five frames would correspond to a common interpolation level. For example, in frames 6 through 10, the target object (e.g., a baseball) would be presented at the 52.5% interpolation level and might appear in frame 9; as such, frames 6, 7, 8, and 10 would therefore depict various distractor objects (e.g., a house, a dog, etc.), also at the 52.5% interpolation level. For each of the five 75-frame RISE sequences, all the target and distractor images were interpolated toward a single, common random phase spectrum. This results in the images becoming increasing indistinguishable the more they are degraded.

2.3.2.3. Procedure. The experiment consisted of five blocks, each corresponding to one of the target objects and consisting of one mixed RISE onset subsequence followed by its complementary offset subsequence. The stimuli were presented on a gamma-corrected CRT display (minimum luminance: $\sim$0–5 cd/m^2; maximum: $\sim$90 cd/m^2) in a dimly lit room ($\sim$5–10 cd/m^2) and were viewed binocularly. Each image in these sequences was presented for 750 ms and subtended approximately 8° of viewing angle. At the end of each 750 ms presentation, the image frame was overwritten with a black square. During this self-timed inter-stimulus period, subjects pressed one of two keys, indicating whether they had or had not recognized the object in the frame just presented. Subjects were naive as to what images would appear in each RISE sequence and naive also, during onset subsequences, as to which object was considered to be the "target." As such, during onset subsequences, when subjects first reported being able to recognize each of the objects, target or distractor, we were able to confirm their reports by

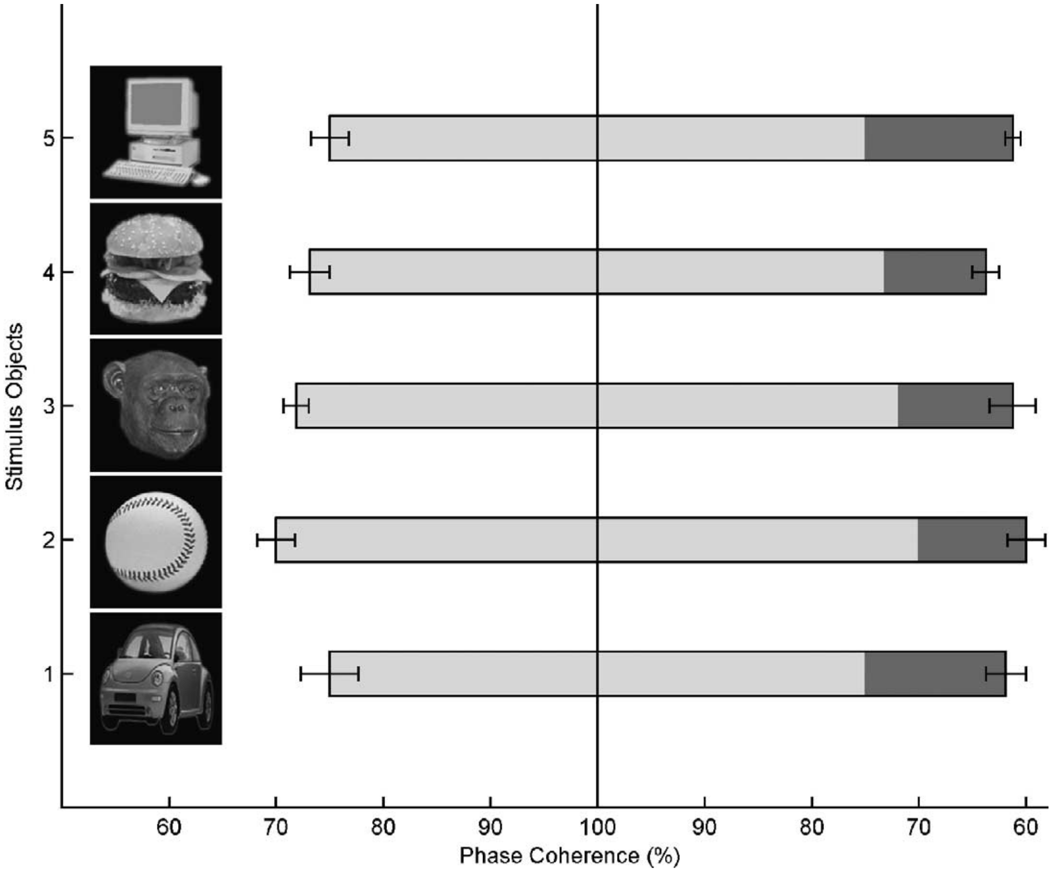

Fig. 7. Onset and offset of object perception in RISE sequences for five objects, represented by the left and right edges of each bar, respectively. Observers exhibit a marked perceptual hysteresis during the offset subsequence—the transition from light to dark gray in each bar serves to indicate the reflection of the onset point about the vertical axis. Data are averaged across four observers.

simply requiring that they also explicitly identify the object seen (by typing a word or two). At the beginning of each offset subsequence, the target object was explicitly singled out and perceptual offset was measured using the distractor technique described above.

2.3.3. Results

Fig. 7 shows the onset and offset points for the five RISE sequences averaged across the four observers. It is interesting to observe that although the progression of the RISE transformations is relatively gradual, for all of the objects there appears to be fairly good agreement (i.e., relatively low variance) across subjects for both the onset and especially offset measures. Coincidentally, the subjects' responses included very few false alarms: over all the images presented to all the subjects, there was only one incorrect (i.e., premature and subsequently corrected) recognition reported during the onset subsequences, and, during the most degraded tail of the offset subsequences, there were in total only four instances when subjects reported seeing the target image when it had not in fact been presented. (In general, of course, false alarms can be a useful tool for analyzing the subjects' percepts and behavioral biases, and it is conceivable that either the RISE images themselves or the instructions to the subjects could be modified in such a way as to promoted more false alarms. For our current purposes, however, it seems beneficial and encouraging that there were so few incorrect responses.) These findings correspond well with the participants' subjective reports of the perceptual onset and offset transitions as being fairly "sharp," and that they were neither guessing nor struggling to gauge their level of confidence before providing each response. Taken together, these results demonstrate well the ability of the RISE technique to objectively and quantitatively assess perceptual onsets and offsets, and to do so in a somewhat natural manner that agrees well with the viewers' subjective experiences.

Before concluding the discussion of the results of this experiment, it is of particular interest to note the significant amount of perceptual hysteresis observed in the RISE sequences for all objects. That is, offset can be seen to occur at a level of image degradation much greater than which supports the onset of recognition, $F(1, 30) = 101.4$, $p \ll .001$. This result and its implications are discussed further below.

3. Applications of the RISE paradigm

The RISE paradigm can be a very flexible and powerful tool in the investigation of several open questions in high-level vision. In this section, we explore some of the more important and intriguing applications of RISE, discussing a variety of RISE experiments that have already been conducted as well as a number of exciting possibilities for the future.

3.1. Characterizing the neural substrates of object perception

The characterization of the neural substrates of object perception is an undertaking of profound significance in psychology and neuroscience. Progress on this front not only brings us closer to understanding the functional architecture of the brain, it may also bear more tangible benefits, such as improved diagnostic and therapeutic approaches in the clinical domain. Previous studies have demonstrated the role of various parts of the brain in the processing of basic perceptual attributes such as motion and color (Newsome & Pare, 1988; Van Essen & DeYoe, 1995; Zeki et al., 1991). However, in terms of the neural substrates of more complex perceptual faculties, though much progress has been made over the past several years (e.g., Martin, Wiggs, Ungerleider, & Haxby, 1996; Perrett et al., 1992; Puce et al., 1995), significant gaps remain in our understanding.

Consider, for example, the well-researched domain of human face perception. In a typical brain imaging or electrophysiology study, a neuron/area that responds more to images of faces than to non-face distractors is often considered to be a "face-cell/area" (Perrett et al., 1992; Puce et al., 1995). However, this methodology may not convincingly establish that the neural response is indeed correlated with the "faceness" of the stimuli. The differential neural response could very well be driven by some other attribute of the stimuli that has little to do with their being (or not being) images of faces. In other words, it is conceivable that neurons or brain regions that appear to be differentially responsive to images of faces (i.e., as opposed to images of non-face objects) nevertheless might also be responsive to images that are, in some important way, "near" face images in the space of images but are not otherwise subjectively perceived as faces or even face-like. Kobatake and Tanaka's experiments (1994), wherein complex visual stimuli, such as hands and tiger heads, were "simplified" without decrement in neuronal responses, are a case in point.

The progressive change in neural activity, and how it correlates with changes in conscious perceptual responses, as stimuli are systematically and continuously varied can be much more diagnostic in establishing links between perceptual processes and neural activity. To this end, RISE explores not merely the absolute levels of perceptual/neuronal responses for individual images, but also how responses change as one moves towards or away from these images in a systematic fashion along continuous trajectories in image space. Along these trajectories, covariance of behavioral responses with the profile of neural activity, if found, would strongly implicate a candidate neural substrate in the high-level perceptual processing of these stimuli.

For example, evidence of hysteresis in the neural response, correlated with perceptual hysteresis during the offset RISE subsequence could be particularly important for determining whether the measured neural responses are purely stimulus driven or are related to the object percept. Such investigations could be conducted using a combination of RISE and functional imaging or electrophysiological techniques and would involve correlating neural recordings with concurrently obtained behavioral data from either human or non-human subjects. It is important to restate the fact that images in RISE offset subsequences are literally identical to their counterparts in the onset subsequences, thus allowing for simple and direct comparisons of their corresponding neural responses.

This simple strategy is similar to that employed in recent functional magnetic resonance imaging (fMRI) work by Kleinschmidt, Buchel, Hutton, Friston, and Frackowiak (2002) studying letter recognition, wherein subjects' perceptual reports were recorded as the luminance contrast of a noisy image of a letter was ramped up and then down. (See also Wilson, 1977 for earlier studies and modeling of hysteresis in binocular grating perception.) Because subjective reports of perceptual offset occurred at lower levels of contrast than those corresponding to perceptual onset, the authors could compare the functional imaging data for identical images that had resulted in different perceptual reports. However, a similar analysis of the change in neural activity at the pre-onset to onset transition (as well as the offset/post-offset transition) would be problematic with this approach, given that the perceptual events of interest are confounded with the increase in stimulus contrast required to elicit them—a direct result of using contrast ramping, rather than another degradation technique which could control for changes in contrast, as the method for driving the perceptual changes of interest.

It is worth noting that while higher level perceptual processes such as object and face recognition are generally considered to be relatively invariant to changes in luminance contrast, this clearly does not hold within the range of contrast corresponding to the perceptual thresholds themselves. (Otherwise, the manipulation by Kleinschmidt et al., 2002, for example, simply would not work in the first place.) Moreover, even when considering the recognition of object images presented at contrast levels above perceptual threshold, a recent fMRI study of the corresponding brain activity along the visual cortical pathway has shown that the notion of contrast invariance in higher level processing should be considered more a matter of degrees than an absolute (Avidan et al., 2002). During changes in luminance contrast, activity in early visual cortex (e.g., V1) was seen to vary more than the activity in the lateral occipital complex (LOC), again supporting the notion of increasing contrast invariance at higher levels of visual processing; however, it is important to not discount the finding that the responses measured in the LOC nevertheless did also vary with changes in contrast. (To further complicate matters, the degree of contrast-sensitivity in the higher visual areas was also modulated by the nature of the objects presented (e.g., faces vs. cars).) As such, even though activity in higher level visual areas such as the LOC is generally correlated with relatively complex visual processes (e.g., object shape processing; Kourtzi & Kanwisher, 2001), this activity can also be modulated, in a non-trivial manner, by lower level features of the stimulus such as contrast.

A nice illustration of the integration of RISE with neuroimaging techniques may be seen in a recent face perception experiment by Liu, Harris, and Kanwisher (2002) using magnetoencephalography (MEG). These authors used a variant of RISE to produce degraded images of faces and houses and to determine thresholds for, and characterize the neural responses correlated with, successful versus unsuccessful between- and within-class discriminations. With this approach, it was possible to isolate a very early (100 ms) occipitotemporal MEG signal component ("M100") that appears to be tied to the successful categorization of a stimulus as a face. Notably, unlike the face-selective N200/M170 component previously characterized by Allison, Puce, Spencer, and McCarthy (1999) and others, this earlier M100 component does not seem to be correlated specifically with successful identification of individual faces.

Again, it is of particular interest here to note one's ability, using such an approach, to compare early neural responses during trials in which subjects did or did not experience a task-relevant object percept (e.g., were or were not able to make within- and between-class discriminations of faces and houses). Because of the controls placed on the image processing, neural responses for each of these images could be readily compared to one another with little concern of the confounding contributions of differences in a number of important low-level image properties. Also, while differences in appearance necessarily exist between the images corresponding to the thresholds for between- versus within-class discriminations in this particular task, the crucial difference between the two conditions is in the perceptual state (and behavioral response) of the subject. (Indeed, while subjects generally tend to experience "eureka"-like perceptual onsets with RISE, the image transform itself can progress rather gradually. As such, in addition to their strictly controlled low-level properties, objectively speaking the images on either side of the perceptual onset and offset thresholds may, in fact, also be rather similar in appearance.) This is not unlike the ability to directly compare neural responses for correct versus incorrect trials for repeated presentations of images at a given threshold level of degradation (i.e., images that are actually identical but which elicit different neural and perceptual responses at different times).

Along these lines, we have also undertaken some simple RISE-based MEG experiments of object and face perception (Sadr & Sinha, 2003; Sadr, 2003), with plans for fMRI experiments in the near future. In these experiments, neural activity is recorded using MEG as subjects view and respond to RISE sequences of objects and faces. As above, the analysis of the neural signals is guided by subjects' perceptual reports: with behavioral responses in hand, RISE stimulus trials and corresponding MEG data are classified *post hoc* into trials corresponding to (as well as preceding and following) perceptual onset and offset. As with Liu et al. (2002), we are able to identify specific occipitotemporal MEG signal components corresponding to the conscious perception of faces and objects. Further, by exploiting the basic presentation protocol of RISE onset and offset sequences (and by presenting the RISE sequence of each object or face only once), we are able to characterize the changes in neural responses coincident with perceptual onset and offset, along with those related to the phenomenon of perceptual hysteresis.

Analogous fMRI experiments could greatly advance our ability to characterize the spatial patterns of neural activity associated with perceptual onset and might allow us to better distinguish the activity correlated with the visual perception of objects of different classes. Such experiments, and simple variants thereof, may help further our understanding of the neural substrates underlying a number of important aspects of conscious perception (e.g., categorical perception) as well as such phenomena as perceptual learning, priming, and hysteresis. Moreover, such work would allow a more direct integration and comparison of RISE-based techniques with the existing fMRI literature (e.g., Avidan et al., 2002; Kleinschmidt et al., 2002; and others).

3.2. Quantitative assessment of priming

Traditionally, the most commonly used indices of priming have been the reduction of response latencies or the improvement of other task-relevant measures of performance (e.g., Bartram, 1974; Kosslyn, 1994). The RISE protocol provides a new priming index: the position along the pattern evolution axis where an observer first recognizes the object being displayed. This is a measure of the minimum amount of visual information a subject needs to perform the detection task. Similar thinking underlies the line-drawing fragmentation (Snodgrass & Feenan, 1990) and gradual "unmasking" (James et al., 2000) paradigms also used in the study of visual priming, particularly repetition priming. In fact, it is important to point out the relationship of these techniques to a seminal experiment by Bruner and Potter (1964) in which subjects formed early, and invariably incorrect, hypotheses regarding the content of very blurry images; primed in this way, subjects took much longer to recognize the objects depicted in these images as they were brought progressively into focus. In our case, we control a number of important image properties that can be confounded with higher level perceptual effects during the progression of the image sequences. Also, we have decided here to use our technique to study a form of priming not based on repeated visual presentations of the target object images. That is, we prime subjects while still leaving them naive as to the appearance and, in fact, the identity of the visual target stimulus. The simple RISE experiment described below demonstrates that priming decreases the amount of information required to visually perceive an object, leading to a shift toward earlier onsets along the RISE trajectory.

3.2.1. Experiment B: priming perceptual onset

3.2.1.1. Aim. The purpose of this experiment is to illustrate the use of RISE in studying the degree to which non-visual priming may shift the point of perceptual onset during the evolution of an object from a seemingly random image.

3.2.1.2. Methods.

Participants. Eight MIT students participated in this study. All subjects provided informed consent and received payment according to MIT guidelines.

Stimuli. The stimuli for the priming experiment were identical to those used in Experiment A.

Procedure. The procedure for the priming experiment was entirely identical to that of Experiment A except for one manipulation, a concurrent word memorization task. Prior to each RISE onset subsequence, the subject was asked to commit to memory one word and instructed that their recall for that word would be tested at the end of the image sequence. They were not told that for approximately half the trials, the word presented matched the common name for the target image in the sequence (e.g., “baseball,” or “car”); for the other trials, this word was unrelated to the target image that would be seen. The assignment of matching versus non-matching trials was counter-balanced across subjects.

3.2.1.3. Results. Fig. 8 shows the onset points for the five RISE sequences, averaged across matching versus non-matching verbal prime conditions. It can be seen that when the concurrent memory task involved a matching word, there was a significant shift of the perceptual onset point toward a more degraded level, $F(1, 30) = 34.34$, $p \ll .001$. This suggests that the matching verbal prime manipulation resulted in a reduction of the amount of visual information required by subjects to correctly recognize the target images, even though the subjects were naive to both the identity and appearance of the objects that would be seen. There was no significant effect of or interaction with the specific identity of the object for each trial. Also, although it may not be entirely appropriate to compare data across experiments, it is perhaps worth noting that there was no significant difference between the onset points in the unrelated-prime trials and the corresponding onset points measured in Experiment A (no priming), but there was a significant difference between the magnitude of this priming effect and that of the hysteresis effect seen in Experiment A, $F(1, 30) = 26.2$, $p \ll .001$—the hysteresis effect seems to produce a greater shift in the perceptual threshold.

The assessment of priming using RISE is particularly convenient because it does not require precise measurement of small temporal effects or carefully controlled tachistoscopic image presentations. If necessary, one could even perform RISE experiments, priming or otherwise, without a computer; in the clinical setting, for example, patients could be tested using a set of pre-printed cards. (One would first calibrate the printing process, of course, so as to not disrupt the important image properties controlled by RISE in the digital images.) It is also of critical importance that RISE allows priming and other perceptual effects to be studied independently of motor influences. As long as the subject can in some way report the occurrence of these perceptual events, the specific details of the motor responses are not necessarily of concern. One consequence of this is that higher level perceptual processing (and deficits thereof) can be studied even with subjects with motor impairments and/or in settings and situations in which motor responses, especially reaction times, cannot be recorded precisely or reliably. Further, because RISE transformations can be set to progress at any desired rate, this technique affords

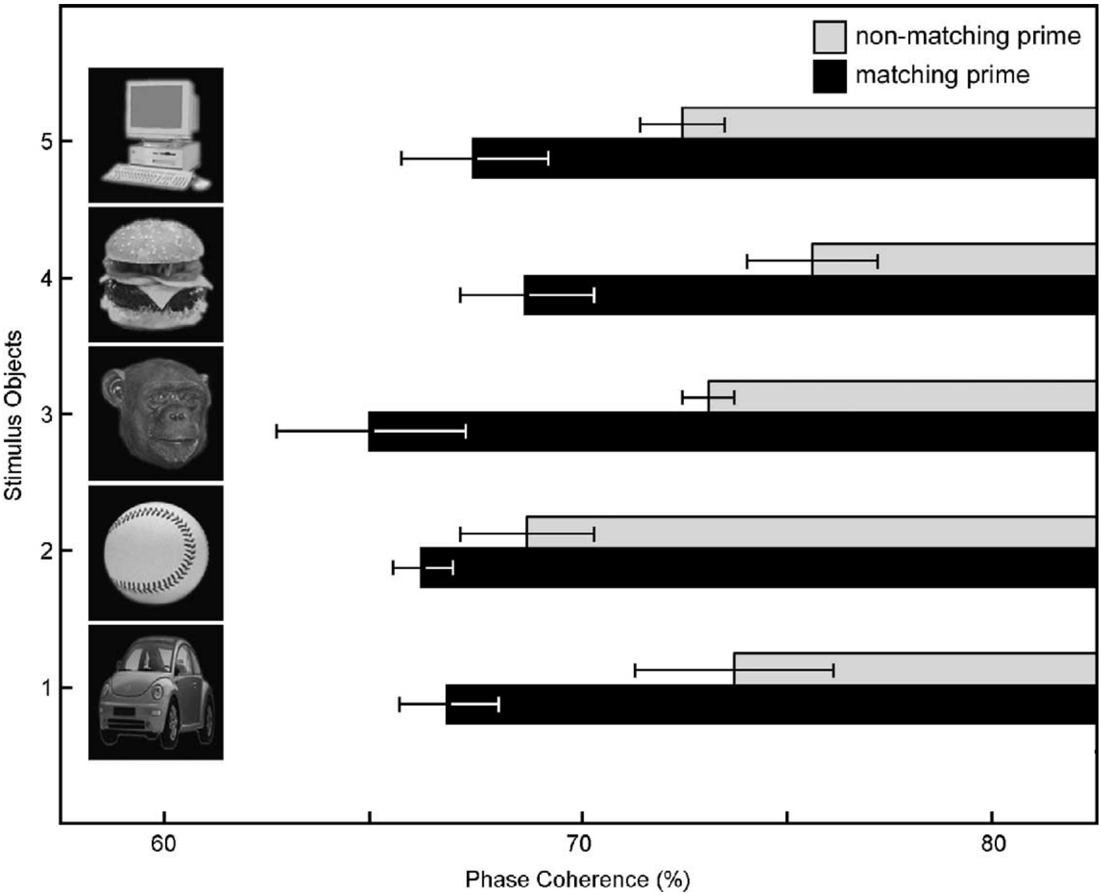

Fig. 8. Comparing onset of object perception with matching versus unrelated verbal priming induced by a concurrent word memory task. Black bars correspond to mixed RISE presentations verbally primed by a word matching the target objects' name, while light gray bars correspond to the same RISE sequences primed by an unrelated word. Matching word primes resulted in a significant shift toward earlier recognition of the target. Data are averaged across four observers per condition.

almost arbitrarily high sensitivity to the subtle effects of various priming manipulations. In a sense, this is analogous to having the ability to dilate time in a reaction time experiment.

3.3. Developmental and clinical studies

3.3.1. Normal and abnormal perceptual learning and development

Just as for experiments in priming, shifts in the onset point may be a useful measure in studies of perceptual learning and development. For instance, one could use RISE to study the development of children's object encoding strategies, thought to progress from being local feature-based to being more holistic or configural (Carey & Diamond, 1977, 1994). One can generate and present to children RISE sequences in which configural information becomes evident sooner than fine featural details; we would hypothesize that, over time, children's onset points will migrate out from the fully formed image. It could be instructive to correlate this migration with other indices of configural coding, such as recognition performance with inverted faces (Bartlett & Searcy, 1993; Brooks & Goldstein, 1963; Diamond & Carey, 1986). Such an approach would bear some relation to those taken by Gollin (1960) and others (e.g., French et al., 2002) in their studies and modeling of perceptual and cognitive development and learning.

It is worth noting that data regarding onset points in normal child populations can also serve as references against which to study the perceptual development of children with differing, even abnormal developmental histories. Differences in the perceptual onset point for a given child relative to that of age-matched controls can be used to detect developmental problems ranging from the purely visual to those more cognitive in nature. The effectiveness of RISE as a sensitive tool for discerning the consequences of atypical developmental histories has already been borne out in a recent study which, using RISE sequences of faces as its stimuli, found an enhanced perceptual sensitivity of physically abused children to anger cues in faces, along with a reduced sensitivity to sadness cues (Pollak & Sinha, 2002).

3.3.2. *Visual agnosias and prosopagnosia*

In addition to the investigation of developmental changes and abnormalities, RISE is also well suited to the study (and perhaps even the diagnosis) of other high-level perceptual deficits. Associative agnosias and prosopagnosia, disorders in which basic visual processes are spared but object- or face-recognition is specifically compromised (Damasio, Damasio, & van Hoessen, 1982; Warrington, 1982; Farah, 1990; Rumiati et al., 1994), are of particular interest. Detecting and diagnosing visual agnosias can be a difficult undertaking, and current tests, such as the Birmingham University Neuropsychological Screen and Snodgrass and Vanderwart's test set (1980), may not be sufficiently sensitive to detect subtle deficits. The ability to name the objects depicted in the well-formed images typical of such tests does not necessarily guarantee that recognition ability is fully intact, and it is possible that certain perceptual deficiencies might become evident only with systematically degraded stimuli. As such, by providing a quantitative measure of the minimal amount of coherent visual information required by an individual to recognize an image, RISE may serve as a more powerful tool for detecting and diagnosing visual agnosia. In particular, RISE might facilitate the detection of subtle and progressive agnosias or prosopagnosia (e.g., Mendez & Ghajarnia, 2001) at relatively early stages of advancement.

3.4. *Studying top–down influences on early visual areas*

It is reasonable to expect that phenomena such as perceptual hysteresis, of the kind observed in RISE experiments, rely at least in part on high-level visual processes. It would be interesting to determine if and in what manner such processes exert top–down influences on early visual areas (Hupe et al., 2001; Jones, Sinha, Poggio, & Vetter, 1997; Seghier et al., 2000; Sinha & Poggio, 1996, 2002). If, as discussed above, the activity in higher visual areas exhibits hysteresis corresponding to perceptual measures, one could also search for evidence of such hysteresis in earlier visual areas. As a starting point, one could test, for example, whether the firing of an orientation-specific V1 cell is tied strictly to the presence of an oriented edge in an image or if, during a RISE offset subsequence, its firing might survive the degradation of that edge, perhaps even to the point of perceptual offset. A similar experimental design can be used to examine the low-level neural correlates of perceptual priming and learning.

4. Discussion

The intent of the present paper is to provide a working description of the RISE paradigm, to supply some simple illustrations of its experimental use, and to discuss a few of the ways in which it may be used in further explorations of important aspects of high-level visual perception. However, the above exposition of the RISE paradigm is focused on one relatively simple implementation of RISE, and it may be worthwhile to consider a few important variations on this theme, along with a closer look at both high- and low-dimensional approaches to stimulus spaces and trajectories.

In one perhaps obvious extension of RISE, one might imagine generalizing the technique beyond individual, static source images to the domain of dynamic stimuli, such as image sequences. For example, an entire image sequence depicting an object in motion could be systematically subjected to progressive levels of degradation. This would produce an ordered set of image sequences, each of which (varying from fully degraded to pristine) could be presented in turn, just as an ordered set of static degraded images are presented during a simple RISE presentation. Alternatively, the time-course of RISE (evolution and/or degradation) could be arranged to coincide with the time-course of the dynamic event(s) depicted in the image sequence. Conceivably, analogous approaches could also be taken for the manipulation of other time-varying signals, such as speech.

Nevertheless, such technical variations as these, along with the relatively simple illustrations presented throughout this paper, do not highlight a key attribute of the RISE paradigm: the ability to create, for a single source image, multiple image sequences depicting transformations to and from numerous end-points, random or otherwise. Returning to the notion of images as points in a multidimensional image space, the basic idea would be to approach a given image of interest from not one but several paths and, for each path, to determine the points of perceptual onset and offset. In effect, onset and offset points would be recast instead as surfaces, and it is reasonable to expect that these surfaces may not have simple (e.g., spherical) geometries. In this manner, using finely-sampled trajectories to and from numerous, diverse end-points, one should be able to describe the complex structure of these perceptual thresholds, as well as the effects of such phenomena as priming and neurological deficits on these threshold surfaces. (That said, in a recent replication of Experiments A and B, using image sequences based on the same objects but generated using different random seeds, it was nevertheless reassuring to find no significant differences in the overall results described above.)

In a simple multi-trajectory RISE experiment, one may prefer at the outset to measure the multiple perceptual thresholds across rather than within subjects, since, for a given subject, repeated exposure to one object would be expected to result in increasingly priming-shifted onset measures. However, within-subject designs are still feasible, particularly if a number of different objects are tested in a randomly interleaved manner across numerous blocks. The data can be depicted using a combination of schematic polar-like plots, along with visualizations of the threshold images themselves—for example, the image at the mean onset point for one trajectory could be compared with that for any number of other trajectories, as well as with the original source image itself, in order to represent the visual information most likely involved in the recognition of a given object.

It so happens, however, that with this approach, as described above in its simplest terms, it may be difficult to extract object-parts-based analyses of the visual information driving recognition. One may contrast this with various techniques designed precisely and primarily for the extraction of task-relevant image structure (e.g., Ahumada, 1987; Mangini & Biederman, 2001; Murray, Bennett, & Sekuler, 2002). This is partly due to our having thus far described only a global-image implementation of RISE (which, further, manipulates the full phase spectrum; see below), and while it is true that across a set of differing random trajectories there will be a differential evolution of the parts of the source object, it is also true that in each RISE sequence the images evolve and degrade globally. A comparison of pre- versus post-onset images, for example, may tend to reveal that passing through the onset threshold coincides with a global improvement in the appearance of the image (particularly the more salient, e.g., higher contrast, parts of the image) rather than with the revelation of a few particularly important (and supposedly diagnostic) image regions. One solution to this apparent limitation is simply to selectively apply the RISE transforms to sub-regions of the source image. For example, one could alternatively evolve and degrade different parts of a face (e.g., eyebrows, mouth, etc.) rather than the whole face (Pollak & Sinha, 2002) in order to assess the relative contributions of these face parts in the expression and visual analysis of various emotions.

It is interesting, then, that to further expand the functionality of the basic RISE image processing technique, one of the simplest and most obvious embellishments also serves to increase its resemblance to a number of other techniques that have focused specifically on local, parts-based image manipulations in order to study the representations underlying visual recognition (e.g., Gosselin & Schyns, 2001; Pomplun, Ritter, & Velichkovsky, 1996). In addition, just as one can perform a spatially local version of RISE, one could also perform RISE phase manipulations that are not global in the frequency domain. That is, one could progressively manipulate only the phase values corresponding to certain spatial frequencies of interest in order to study their relative involvement in a given visual task (and/or their relative contribution to the appearance of a given object or object class). This technique is essentially what was used by Nasanen (1999) to create individual face images to investigate the relative importance of different frequency bands for face recognition, and it has been more recently applied to the study of basic-level categorization (Schyns & Gosselin, 2002).

Even with the simplest implementation of RISE, however, a full appreciation of the workings and applications of the technique would benefit from a good intuition of the nature of the image trajectories involved, not to mention the space in which they reside. Perhaps the best descriptions of these are also the most frank. Within the full space of all possible images, all images produced and presented in RISE reside in a subspace (or manifold) characterized by a shared global power spectrum, luminance, and contrast; RISE trajectories are smooth paths that reside in this space and connect pairs of selected images (e.g., an object image and a randomized counterpart) via a direct linear interpolation of their respective Fourier phase coordinates. One could even say that there is a bias in the sampling of these paths—that is, a bias of selecting only the image trajectories that control for power spectra and luminance, yield monotonic evolution/degradation of the object image, directly and smoothly connect the end-point images, etc. Needless to say, we think such constraints are well-founded for our current purposes, and there are no further criteria by which images are included or excluded or by which such image space excursions are sampled. Indeed, as discussed above, one may even exploit the existence of many possible trajectories to/from a given object image in quantifying perceptual threshold surfaces, all while controlling a number of important image properties.

It is also worthwhile to relate the general basis of our approach with other work based on the exploration of lower dimensional, parameterized shape/feature spaces. A nice illustration of such an approach can be seen in recent work by Leopold, O'Toole, Vetter, and Blanz (2001) employing morphs (e.g., averages, caricatures, and anti-caricatures) of high-resolution 3-D models of laser-scanned faces. Certainly, the exploration of a relatively low-dimensional space with concomitant perceptual measures (e.g., face identification or gender discrimination) may be quite expedient, and often the correspondence between the parameterized space and the associated image properties is relatively transparent (e.g., values along one or more dimensions might represent, say, different nose lengths). Further, another low-dimensional shape-based approach has been central to the development of a very intriguing, formalized representation of shape similarity and its application to 3-D object discrimination (Edelman, 1995). In comparison, the higher dimensional photometric approach taken in RISE may be considered agnostic, in a sense, to the underlying, lower dimensional object structures that contribute to the appearance of the object images of interest. Consequently, however, RISE may be applied to the manipulation and use of a vast and varied set of source images. As a result, RISE may be in certain respects more suitable for the study of perceptual and neural responses to actual images of arbitrary, real-world objects/scenes of interest when, for example, 3-D renderings of parameterized, low-dimensional models are either unavailable, unfeasible, or perhaps in some way inappropriate for a particular experimental objective.

At this point it may be appropriate, in fact, to briefly revisit the motivations that have shaped the development of RISE in its current form. A quick perusal of the experimental and theoretical discussions above may rightly impress upon the reader that while RISE may be useful in exploring the influence of image structure on high-level visual processes, it is in many ways a paradigm that is aimed more toward providing a simple and effective procedure by which to manipulate and study the visual processes themselves. That is to say, it is simply more the flavor of RISE, in its current form, to empower an experimenter to elicit key perceptual transitions and thereby study their neural correlates and/or their hysteresis, susceptibility to priming, change during childhood development, etc.—and to do so while eliminating some important low-level confounds intrinsic to a number of other common techniques—rather than to directly quantify, for example, the relative importance of certain image regions or spatial frequencies in the performance of a given recognition task (e.g., the importance of the eye region for facial identity but apparently not for expressiveness, the use of spatial frequencies between 11.25 and 22.5 cycles for identification vs. 5.62–11.25 cycles for expressiveness judgments, and so on; Schyns et al., 2002).

In summary, we hope this paper shows the relative strengths and flexibility of the RISE paradigm for the exploration of a number of important issues in high-level vision. In its simplest form, RISE can be used to collect information about the formation and disruption of object percepts, but it can very easily also serve to sensitively measure the perceptual effects of priming, to quantitatively and objectively study perceptual hysteresis, and to examine the consequences of abnormal perceptual development and learning. It represents a new approach to the study of the neural substrates of high-level visual perception, and it could also find use in the clinical setting to assess and perhaps diagnose certain visual disorders. Despite its versatility, it is a computationally straightforward technique that is relatively easy to implement. In fact, once RISE image sequences are created, their experimental use does not necessarily require even the use of a computer—the image frames could conceivably be presented in printed form (e.g., in a flip-book) and subjects' responses can be recorded by hand and in a leisurely fashion. RISE derives its power from the simple idea that the investigation of high-level visual perception via behavioral and neural measures can be made more compelling, not to mention easier to interpret, when experiments more thoroughly and systematically explore the space of images.

Acknowledgments

This research was supported in part by an Alfred P. Sloan Fellowship in Neuroscience to Pawan Sinha and by a Howard Hughes Medical Institute Pre-Doctoral Fellowship to Javid Sadr. The authors wish to thank Philippe Schyns and the anonymous reviewers, as well as Patrick J. Bennett, Zoe Kourtzi, Sayan Mukherjee, Aude Oliva, Gonzalo Ramos, Bosco Tjan, Antonio Torralba, and Tabitha Spagnolo Sadr.

References

Ahumada, A. J., Jr. (1987). Putting the visual system noise back in the picture. *Journal of the Optical Society of America A, 4*(12), 2372–2378.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I. Potentials generated in the occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex, 9*, 415–430.

Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., & Malach, R. (2002). Contrast sensitivity in human visual areas and its relationship to object recognition. *Journal of Neurophysiology, 87*, 3102–3116.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*(3), 281–316.

Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology, 6*(3), 325–356.

Benson, P. J., & Perrett, D. I. (1993). Extracting prototypical facial images from exemplars. *Perception, 22*, 257–262.

Bracewell, R. N. (2000). *The Fourier transform and its applications*. Boston: McGraw Hill.

Brooks, R. M., & Goldstein, A. G. (1963). Recognition by children of inverted photographs of faces. *Child Development, 34*, 1033–1040.

Bruner, J. S., & Potter, M. C. (1964). Interference in visual recognition. *Science, 144*, 424–425.

Busey, T. A. (1998). Physical and psychological representation of faces: Evidence from morphing. *Psychological Science, 9*, 476–483.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195*, 312–314.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 213*, 253–274.

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception and Psychophysics, 58*(4), 602–612.

Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology, 12*(14), R476–R477.

Damasio, A., Damasio, H., & van Hoessen, G. (1982). Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology, 32*, 331–341.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, *69*, 1091–1117.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*, 2051–2062.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision.* New York: Oxford University Press.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107–117.

Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R. S. J., et al. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature*, *389*, 596–599.

Edelman, S. (1995). Representation of similarity in 3D object discrimination. *Neural Computation*, *7*, 407–422.

Farah, M. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision.* Cambridge, MA: MIT Press.

French, R. M., Mermillod, M., Quinn, P., Chauvin, A., & Mareschal, D. (2002). The importance of starting blurry: Simulating improved basic-level category learning in infants due to weak visual acuity. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society.*

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, *39*, 3537–3560.

Gollin, E. S. (1960). Developmental studies of visual recognition of incomplete objects. *Perceptual and Motor Skills*, *11*, 289–298.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*, 2261–2271.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, *6*, 316–328.

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

Hupe, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology*, *85*(1), 134–145.

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). The effects of visual object priming on brain activation before and after recognition. *Current Biology*, *10*, 1017–1024.

Jones, M., Sinha, P., Poggio, T. A., & Vetter, T. (1997). Top-down learning of low-level vision tasks. *Current Biology*, *7*, 991–994.

Kleinschmidt, A., Buchel, C., Hutton, C., Friston, K. J., & Frackowiak, R. S. J. (2002). The neural structures expressing perceptual hysteresis in visual letter recognition. *Neuron*, *34*, 659–666.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*, 856–857.

Kohler, S., Kapur, S., Moscovitch, M., Winocur, G., & Houle, S. (1995). Dissociation of pathways for object and spatial vision: A PET study in humans. *Neuroreport*, *6*, 1856–1868.

Kosslyn, S. M. (1994). *Image and brain.* Cambridge, MA: MIT Press.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*, 1506–1509.

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading—I. Normal vision. *Vision Research*, *25*(2), 239–252.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*(1), 89–94.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*(9), 910–916.

Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, *42*, 1165–1184.

Mangini, M. C., & Biederman, I. (2001). Differentiating expression, gender, and identity in faces: Comparing normals, the ideal observer, and a prosopagnosic. *Vision Sciences Society Abstracts*, *1*, 330.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category specific knowledge. *Nature*, *379*, 649–652.

Mendez, M. F., & Ghajarnia, M. (2001). Agnosia for familiar faces and odors in a patient with right temporal lobe dysfunction. *Neurology*, *57*, 519–521.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin and Review*, *8*(3), 454–469.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision*, *2*, 79–104.

Nasanen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, *39*, 3824–3833.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, *8*, 2201–2211.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, *69*, 529–541.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and function of cells responsive to faces in the temporal cortex. *Transactions of the Royal Society of London*, *B225*, 23–30.

Pollak, S. D., & Sinha, P. (2002). Enhanced perceptual sensitivity for anger among physically abused children. *Developmental Psychology*, *38*(5), 784–791.

Pomplun, M., Ritter, H., & Velichkovsky, B. (1996). Disambiguating complex visual information: Toward communication of personal views of a scene. *Perception*, *25*(8), 931–948.

Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extra-striate cortex studied by functional MRI. *Journal of Neurophysiology*, *74*, 1192–1199.

Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Current Biology*, *11*, 846–854.

Rainer, G., & Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, *27*, 179–189.

Rumiati, R. I., Humphreys, G. W., Riddoch, M. J., & Bateman, A. (1994). Visual object agnosia without prosopagnosia or alexia: Evidence for hierarchical theories of visual recognition. *Visual Cognition*, *2/3*, 181–225.

Sadr, J., & Sinha, P. (2001a). *Exploring object perception with random image structure evolution*. Massachusetts Institute of Technology, Artificial Intelligence Laboratory Memo #2001-06.

Sadr, J., & Sinha, P. (2001b). Random image structure evolution (RISE). *Vision Sciences Society Abstracts*, *1*, 83.

Sadr, J., & Sinha, P. (2003). Characterizing object-specific neural correlates of perception. *Vision Sciences Society Abstracts*, *3*, 146.

Sadr, J. (2003). Visual Perception and representation of objects and faces (Doctoral dissertation, Massachusetts Institute of Technology, 2003). *Dissertation Abstracts International*, 64(9B).

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409.

Schyns, P. G., & Gosselin, F. (2002). A natural bias for basic-level object categorizations. *Vision Sciences Society Abstracts*, *2*, 144.

Seghier, M., Dojat, M., Delon-Martin, C., Rubin, C., Warnking, J., Segebarth, C., et al. (2000). Moving illusory contours activate primary visual cortex: An fMRI study. *Cerebral Cortex*, *10*(7), 663–670.

Shelton, C. R. (1998). *3D correspondence.* Masters thesis, MIT Department of Electrical Engineering and Computer Science.

Singh, K. D., Smith, A. T., & Greenlee, M. W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage*, *12*, 550–564.

Sinha, P., & Poggio, T. A. (1996). Role of learning in three-dimensional form perception. *Nature*, *384*, 460–463.

Sinha, P., & Poggio, T. A. (2002). High-level learning of early visual tasks. In M. Fahle & T. Poggio (Eds.), *Perceptual learning*. Cambridge, MA: MIT Press.

Snodgrass, J. G., & Corwin, J. (1988). Perceptual identification thresholds for 150 fragmented pictures from the Snodgrass and Vanderwart picture set. *Perceptual Motor Skills*, *67*, 3–36.

Snodgrass, J. G., & Feenan, K. (1990). Priming effects in picture fragment completion: Support for the perceptual closure hypothesis. *Journal of Experimental Psychology: General*, *119*, 276–296.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 175–215.

Van Essen, D. C., & DeYoe, E. A. (1995). Concurrent processing in primate visual cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 383–400). Cambridge, MA: MIT Press.

Warrington, E. K. (1982). Neuropsychological studies of object recognition. *Philosophical Transactions of the Royal Society of London*, *B298*, 15–33.

Wilson, H. R. (1977). Hysteresis in binocular grating perception: Contrast effects. *Vision Research*, *17*, 843–851.

Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kemard, C., & Frackowiak, R. S. J. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience*, *11*, 641–649.