massachusetts institute of technology — artificial intelligence laboratory

Exploring Object Perception with Random Image Structure Evolution

Javid Sadr and Pawan Sinha

AI Memo 2001-006 March 2001
CBCL Memo 196

© 2001 massachusetts institute of technology, cambridge, ma 02139 usa — www.ai.mit.edu

Abstract

We have developed a technique called RISE (Random Image Structure Evolution), by which one may systematically sample continuous paths in a high-dimensional image space. A basic RISE sequence depicts the evolution of an object's image from a random field, along with the reverse sequence which depicts the transformation of this image back into randomness. The processing steps are designed to ensure that important low-level image attributes such as the frequency spectrum and luminance are held constant throughout a RISE sequence. Experiments based on the RISE paradigm can be used to address some key open issues in object perception. These include determining the neural substrates underlying object perception, the role of prior knowledge and expectation in object perception, and the developmental changes in object perception skills from infancy to adulthood.

This research was sponsored in part by the Alfred P. Sloan Fellowship in neuroscience to Pawan Sinha. Javid Sadr is a Howard Hughes Medical Institute Pre-Doctoral Fellow.

1. Introduction

As depicted schematically in figure 1, images of different objects can be thought of as points lying in a high-dimensional space. Each dimension corresponds to one of the ways in which the image can vary. For instance, a 100x100 pixel image may be represented as a point in a 10,000 dimensional space. Conventional object perception experiments that employ distinct images as stimuli are akin to probing response to isolated points drawn from this space. Examples of such experiments include electrophysiological studies of 'face-neurons' wherein cells' responses are assessed for various objects such as faces, hands, toilet-brushes, etc [Desimone et al, 1984; Perrett et al, 1992]. Functional neuro-imaging studies that attempt to identify object-specific cortical areas follow a similar approach [for instance, Kanwisher et al, 1997]. On the basis of this unsystematic sampling of the image space, cells/areas are declared to be specific or non-specific for particular objects. In essence, a functional form is ascribed to the cells' responses based on sparse sampling over widely separated points. This is an ill-posed undertaking. One way of mitigating this problem is to have dense sampling in the neighborhood of an object point. In other words, we can assess how the response changes as one moves along a continuous trajectory passing through a point of interest in the multidimensional image space.

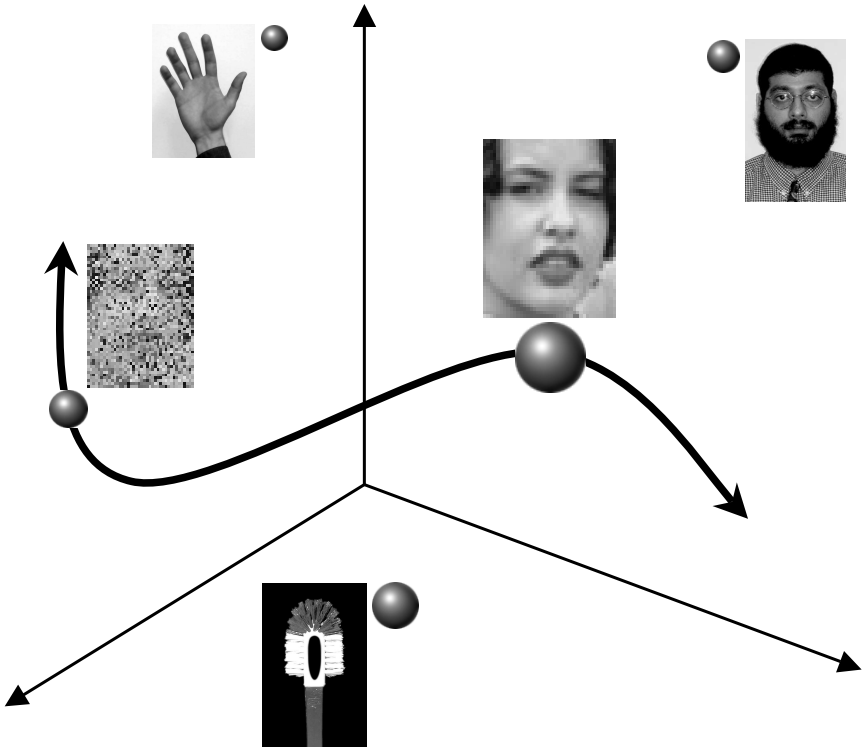

Figure 1. *Conceptually, different points in a high-dimensional 'image space' correspond to different objects, as shown here schematically. Most object perception studies probe behavioral and/or neural responses at randomly selected points in this space. A more systematic, and potentially informative, alternative approach is to examine responses along a continuous trajectory passing through chosen points in this space.*

One can plot changes in a variety of attributes as a function of the trajectory. Besides the perceptual responses for different subjects, these attributes may include theoretical measures of information content in images along the trajectory and measures of neural activity. By analyzing the mutual correlations of the plots across attributes and across subjects, information critical for answering a host of important questions can be obtained. This is the key motivation underlying the RISE (Random Image Structure Evolution) paradigm.

2. The RISE paradigm

2.1. Stimulus image processing

RISE can be thought of as a specific type of morphing procedure [Benson and Perrett, 1993; Busey, 1998]. A very simple version of RISE proceeds by performing pair-wise flips of image regions. As the flips accumulate, the original image dissolves into a random field. The coordinates of the flipped regions are continuously recorded. This allows the sequence to be played backward as well as forward, thus forming a continuous trajectory passing through an intermediate state of the perfect pattern. The first half of a RISE sequence ('onset' subsequence) shows an object emerging from a random field, while the second half ('offset' subsequence) shows the object disappearing back into randomness (see figure 2 for a sample sequence). Thus, the two extremes of a complete RISE sequence are random patterns while the midpoint is a fully constituted object image. The sizes of the regions and the spatial extents of the transpositions (with small extents leading to local structure randomization) are under the experimenter's control.

Besides being computationally simple to implement, this RISE protocol possesses a very attractive characteristic: it precisely maintains important low-level attributes of stimuli, such as overall luminance and color distributions. This avoids contamination of the experimental results with luminance-based low-level artifacts.

However, this implementation of RISE has one drawback in that it does not preserve the frequency spectrum of the source image. Could we overcome this drawback by first equalizing the frequency spectra of two images (say, a random image and an image of interest) and then linearly interpolating between them, as done in the study reported by Rainer and Miller [2000]? Unfortunately, this approach too does not preserve the frequency spectrum throughout the sequence. A simple example illustrates this problem: consider two images showing a vertical sine-wave grating. Assume that the only difference between them is a 180 degree phase shift of the grating. Both images possess the same power spectrum. However, if we were to create intermediate images using linear interpolation, at the half-way point (equal contributions from both images), the resultant image is uniformly gray. This clearly has a different power spectrum compared to the reference images.

An alternative approach is to manipulate the images in the Fourier domain, progressively transforming the phase while holding constant the power spectrum. In the case of the onset portion of a basic RISE sequence, the perfect image evolves from a random-seeming starting image that has been constructed using a random phase matrix combined with the power spectrum of the original image. The onset subsequence is achieved through progressive transformation of the random phase matrix into that of the perfect image, and the offset subsequence is simply this process in reverse (see figure 3 for a sample sequence). In our phase-manipulation implementation of RISE, all images in the sequence have identical power spectra and overall luminance. Further, this manipulation can be performed in ways that ensure monotonic evolution and degradation of the image (e.g., monotonic decrease and increase in the L2 distance from the source image) during the onset and offset subsequences, respectively (figure 4).

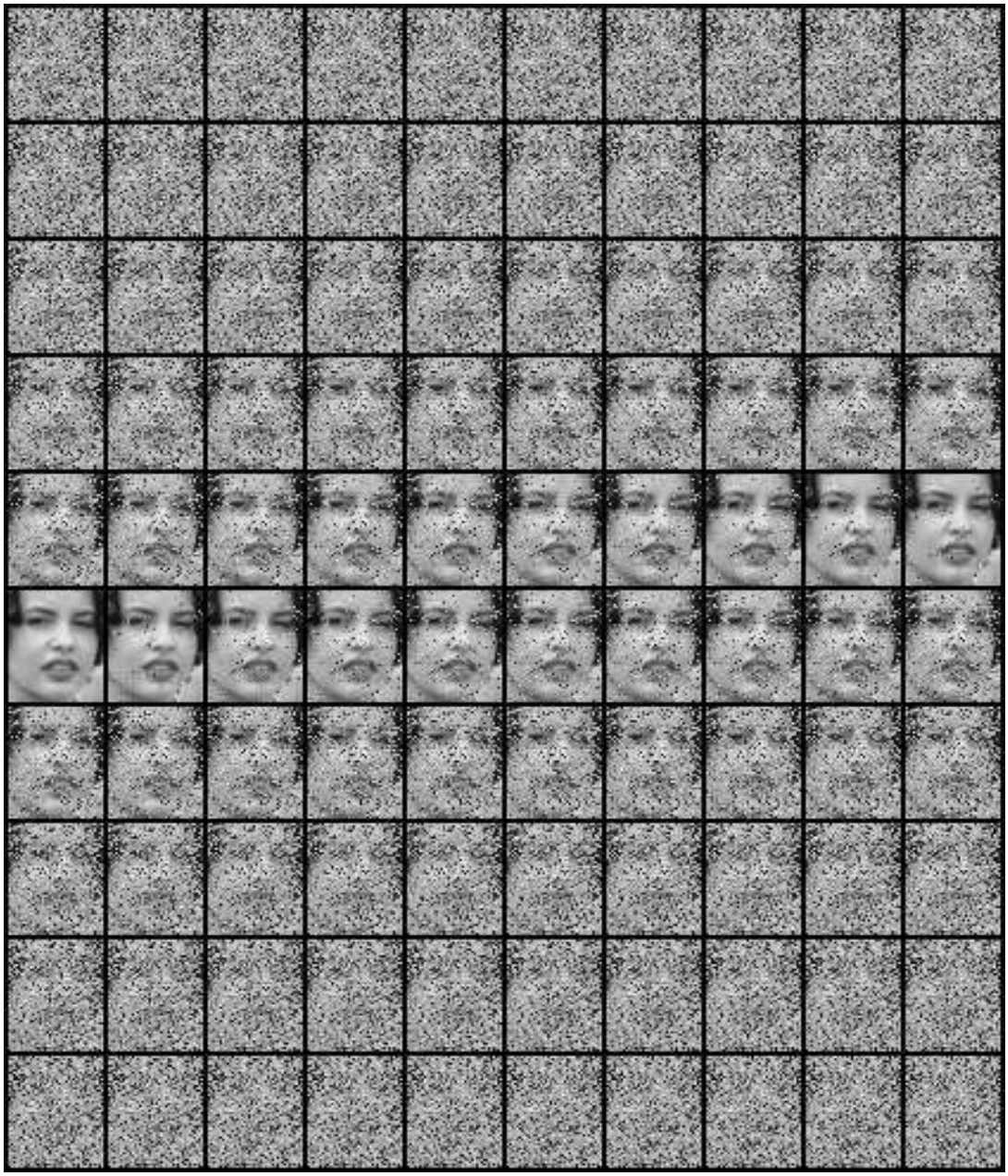

Figure 2. *A sample RISE sequence, generated by pair-wise flips of image regions. A simple presentation of these images would proceed in raster order (i.e., from left to right and top to bottom). The source image appears in the first column of the sixth row.*

Note that in addition to depicting evolution from and degradation to apparently random images, with trivial variation this phase-manipulation technique may also be used to morph between different objects' images. In terms of the high-dimensional image space discussed above, such transformations would correspond to the traversal of trajectories that connect pre-established points (i.e., object images) of interest. As above, the transformation of the phase matrices may be

Figure 3. *A sample RISE sequence generated by progressive degradation of the source image phase matrix.*

brought about by various methods, such as interpolation or random, accumulating substitution of elements. If the source images are first normalized in terms of their power spectra and luminance, these will be held constant throughout the morph sequence, with the only change being in the underlying phase.

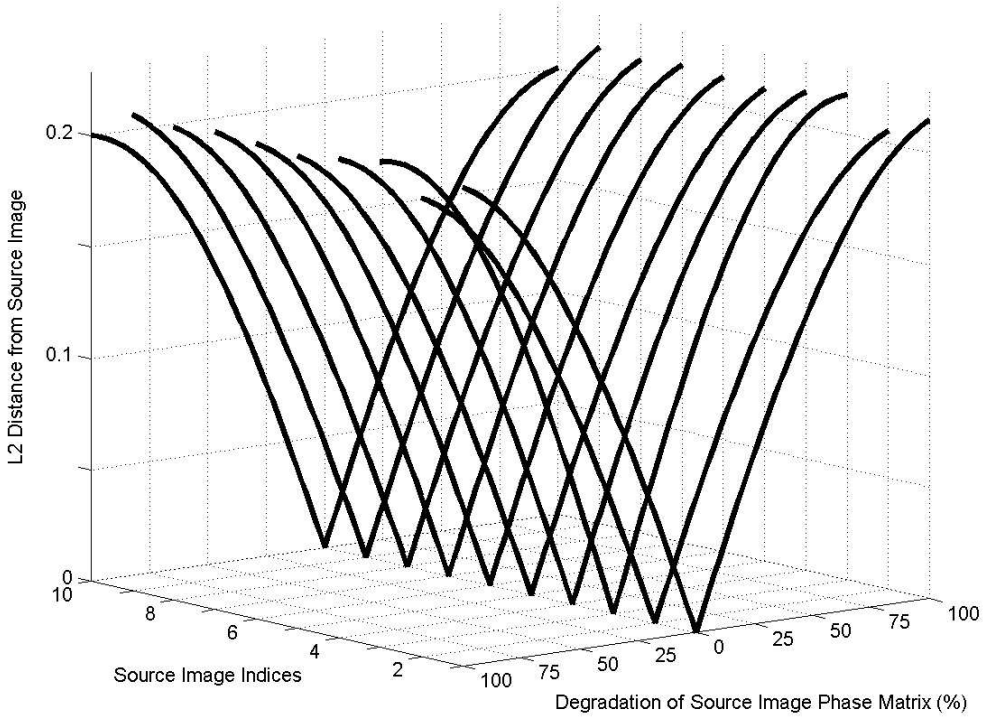

Figure 4. *Ten RISE sequences were generated, each based on one of ten source images. Here are plotted the L2 distances between the original source images and each of the images in its associated RISE sequence. It can be seen that RISE image processing can be performed in such a way as to ensure monotonicity within each of the onset and offset subsequences. (Notice that, as described above, the source images appear in non-degraded form at the midpoint of the RISE sequences, at the transition between the onset and offset subsequences; appropriately, these correspond to L2 distances of 0.) Here, source images were first equalized for luminance and Fourier magnitude.*

Finally, the RISE approach could also be generalized 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, these approaches could also be taken for the manipulation of other time-varying signals, such as speech.

2.2. Basic experimental paradigm

Using the simplest version of a RISE sequence, where just one object emerges from and then dissolves back into a random field, quantitative estimates of two important aspects of an observer's percepts can be obtained.

1. The 'first-detect' or 'onset' point - The position along the initial half of the RISE sequence where the observer is first able to identify the emerging object.

2. The 'last-detect' or 'offset' point - The position along the second half of the RISE sequence beyond which the observer is no longer able to detect the presence of the object in the image.

As discussed in the next section, these two measurements along a RISE sequence can help probe many aspects of object perception. Before proceeding to discuss these potential uses, let us briefly consider how a RISE sequence may actually be presented to observers in an experimental setting.

Observers can view the RISE sequence passively or they may be required to actively indicate object onset and offset points. Passive viewing is better suited to electrophysiology experiments, where the objective is to determine the activation onset and decay for object-specific neurons in anaesthetized or awake but untrained animals. However, for most behavioral experiments, subjects' overt responses will be required to assess their perception of RISE sequences. Subjects can, of course, be asked simply to indicate when they begin and stop seeing the object in the sequence. However, it is desirable to modify this basic idea to obtain objective verification of the observers' reported percepts. We describe two techniques for accomplishing this. There are likely to be many more.

1. Subjects are not told beforehand which object will appear in the RISE sequence. Their task is to identify the object as the onset subsequence progresses. Accuracy of identification serves to validate subjects' verbal reports of object-perception onset.

2. Distractor images are randomly inserted throughout the original RISE offset subsequence (figure 5). These distractors are taken from RISE sequences of other objects' 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 degraded images may or may not be of the original 'target' object. As this mixed offset subsequence proceeds, there will come a point when the subject can no longer reliably detect the presence of the target object's image among the distractors; we take this point to mark the offset of object perception.

Figure 6 shows the onset and offset points for five objects averaged across four observers. It is interesting to note the remarkable amount of hysteresis observed in the RISE sequences for all objects.

3. Potential uses of the RISE paradigm

The RISE paradigm can be helpful in addressing several open issues related to object perception. We list some of the most important ones below.

3.1. What are the neural substrates of object perception?

The characterization of the neural substrates of object perception is an issue of profound significance for all domains of psychology and neuroscience. Progress on this issue not only brings us closer to understanding the functional architecture of the brain, but also has more direct benefits in terms of devising better therapeutic procedures for dealing with brain damage. Previous studies have demonstrated that disparate parts of the brain underlie the processing of

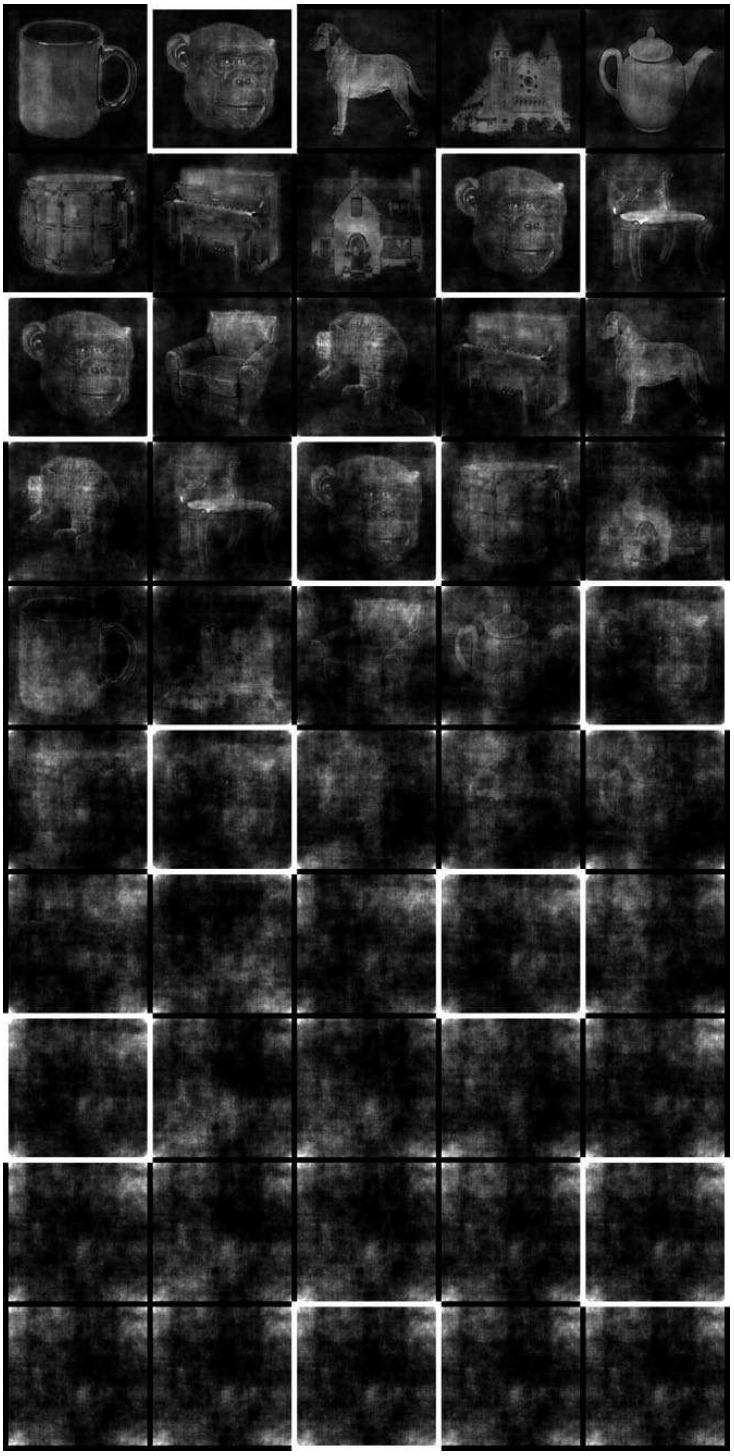

Figure 5. *RISE offset subsequence incorporating distractors. The target object for this sequence is framed in white. Throughout, the images have the same overall luminance and Fourier magnitude. Here, they also gradually converge to a common phase matrix.*

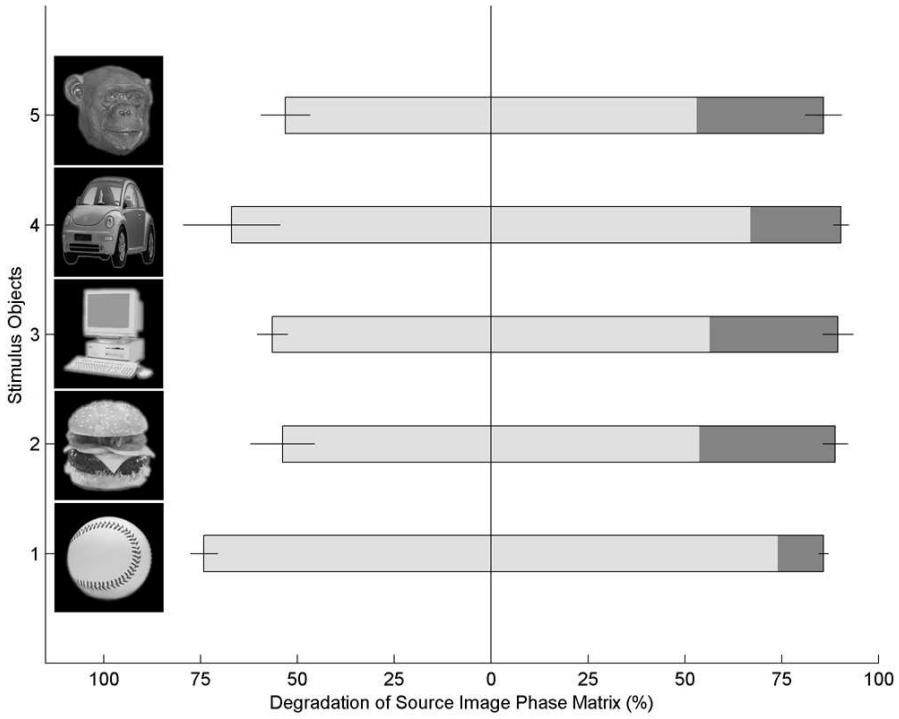

Figure 6. *Onset and offset of object perception in RISE sequences for five objects. Observers exhibit a marked perceptual hysteresis (indicated by dark gray sections in bar graph above) during the offset subsequence. Data are averaged across four observers.*

basic perceptual attributes such as motion and color [Newsome and Pare, 1988; Zeki et al, 1991; Van Essen and DeYoe, 1995]. However, as far as the issue of localization for more complex perceptual entities is concerned, though significant progress has been made over the past several years [Perrett et al, 1992; Martin et al, 1996; Kanwisher et al, 1997; Epstein and Kanwisher, 1998], significant gaps remain in our understanding.

Consider, for instance, the well-researched domain of human faces. In a typical brain imaging or electrophysiology study, an area/neuron that responds more to a face pattern than to non-face distractors is deemed to be a 'face-area / -cell' [Perrett et al, 1992; Kanwisher, 1997]. However, this methodology does not convincingly establish that the neural response is indeed correlated with the 'faceness' of the patterns. It could very well be driven by some other attribute that has little to do with a pattern being a face. Tanaka's recent experiments, wherein he was able to 'simplify' complex patterns without any decrement in neuronal responses, are a case in point [Kobatake and Tanaka, 1994].

The progressive change in the neural activation (and how it correlates with changes in conscious perceptual responses) as the pattern undergoes a continuous transformation along the RISE trajectory can be much more diagnostic in establishing a link between perception and neural response. To this end, RISE explores not merely the absolute levels of perceptual/neuronal responses for individual patterns, but also how the responses change as one moves towards or away from the pattern along a continuous trajectory in the multidimensional space. Covariance of the behavioral response with the neural activation profile, if found, would strongly implicate the

site of the neural activity in the processing of the input pattern. This investigation can be conducted using either functional imaging or electrophysiology techniques. Given the limited temporal resolution of the former, however, the rate of presentation of the frames in the RISE sequence will have to be slowed down. In either case, the experiment will involve correlating neural activation traces with concurrently obtained behavioral data from human or non-human primates. This would allow us to determine whether attributes of conscious perception (such as categorical perception and hysteresis) are observable in neural activation traces. If a high correlation is found between neural and perceptual profiles, it would provide strong evidence in support of the brain-site's involvement in perceptual processing of the stimulus pattern. A negative result, however, would not be very informative. It would merely suggest that regions of the brain besides those that have been investigated so far may be involved in shaping perception of the patterns.

3.2. Can we develop new approaches to the quantitative assessment of object priming?

Traditionally, the most commonly used indices of priming have been reduction in latencies of response or elevation in some measure of performance [Biederman and Cooper, 1991; Cave, 1997; Bar and Biederman, 1998]. The RISE protocol provides a new priming index - the position along the pattern evolution axis where an observer first detects the presence of the pattern. This is a measure of the minimum amount of information a subject needs to perform the detection task. Our pilot data shows that with priming, the amount of information required decreases, leading to a shift of the first-detect point along the RISE evolution axis (figure 7). This measure is particularly convenient because it does not have to rely on precise temporal measurements (priming effects observed in several studies are on the order of a few tens of milliseconds) or carefully controlled tachistoscopic image presentations. The rate of pattern evolution in RISE can be set to any convenient value.

3.3. Can we devise new quantitative measures of perceptual development and learning?

Just as for priming studies, the first-detect point can also be used as a measure of perceptual learning and perceptual development. The kinds of questions that one can ask here include: How does learning affect the position of the first-detect point along the RISE trajectory? Can the position of the first-detect point be used as a quantitative indicator of the different stages in perceptual development? For instance, the RISE paradigm could be used to verify whether children's object encoding strategy progresses from being local feature-based to more holistic or configurational [Carey and Diamond, 1977, 1994]. Given that in a RISE sequence, configurational information becomes evident sooner than fine featural details, we would expect that children's first detect points will migrate out from the fully formed image over time. It would be instructive to correlate this migration with other indices of configural coding, such as recognition performance with inverted faces [Brooks and Goldstein, 1963; Diamond and Carey, 1986; Bartlett and Searcy, 1993]. It would also be interesting to determine if the RISE paradigm can be adapted to serve as a diagnostic test for specific problems in perceptual development.

3.4. Can we devise more sensitive tests for detecting visual agnosias?

It is often very difficult to diagnose the nature of a visual agnosia [Warrington, 1982; Farah, 1990; Rumiati et al, 1994] caused by brain damage. Current tests such as BLTNS (Birmingham University Neuropsychological Screen) and Snodgrass and Vanderwart's test set [1980], involving object naming, are too crude to detect subtle forms of agnosia. An individual's ability to recognize a perfect image does not establish that his/her recognition ability is normal. The deficiencies might become evident with systematically degraded images. The RISE paradigm, by

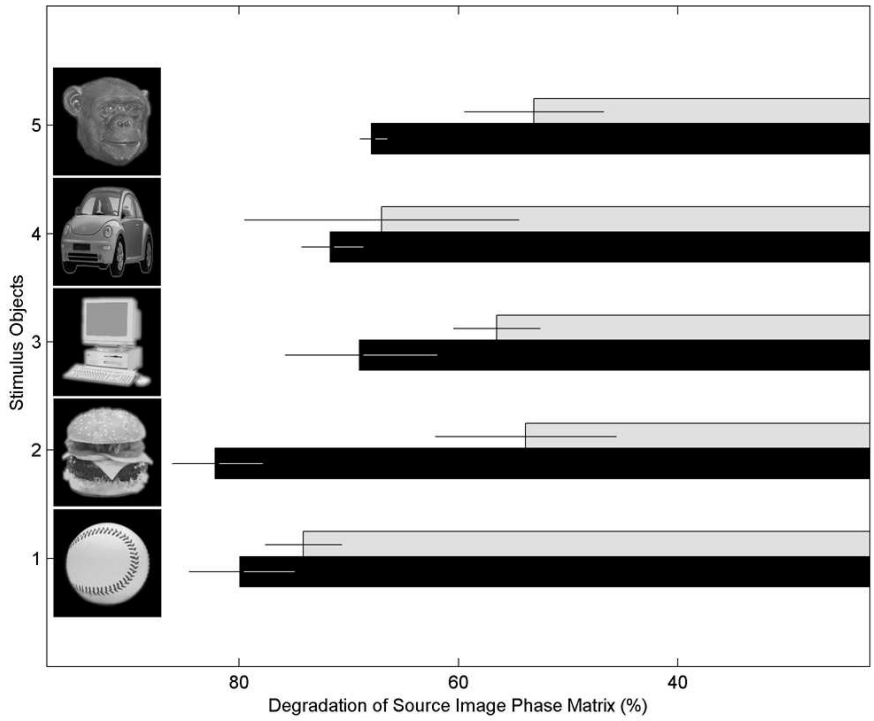

Figure 7*. Comparing onset of object perception with and without verbal priming. Light gray bars correspond to unprimed presentations while black bars correspond to verbally primed presentations. Data are averaged across four observers per condition.*

providing a quantitative measure of the minimal amount of information needed to detect an object, can prove to be much more useful for detecting agnosias. It can also potentially help in distinguishing between different forms of agnosia, such as apperceptive agnosia and dorsal/ ventral simultanagnosia.

3.5. Are early visual areas subject to top-down influences?

Perceptual hysteresis of the kind observed with RISE sequences can be attributed at least partly to high-level visual processing since it requires a maintenance of an object percept. It would be interesting to determine if such processing exerts any top-down influences on the early visual areas [Sinha and Poggio, 1996, 1999; Jones et al, 1997]. Specifically, would the early areas exhibit hysteresis correlated with hysteresis in the higher areas? For instance, would an oriented simple cell continue responding to an edge in an image (presented in a RISE protocol) until the 'offset' point, even after the local structure has been randomized?

3.6. Can we devise a simple scheme for image data encryption and watermarking?

Aside from the research issues listed above, the RISE protocol also has potential practical applications. For instance, it provides a simple way of encrypting image information, with the random seed and evolution extent serving as encryption/decryption keys. With the amount of digital graphical information exploding on the internet, the RISE encryption protocol can make a very timely contribution towards secure transmission of data. Similarly, there is potential for the application of certain aspects of the RISE technique to the domain of digital watermarking.

4. Conclusion

The development of the RISE paradigm presents us with the opportunity to explore a very rich and exciting set of research issues in the area of object perception. In this paper we have focused on the technique and its potential uses, at the expense of experimental data. Experimentation with RISE has recently commenced in our laboratory, and preliminary results bear out the promise of the technique. Additional experiments are underway to address some of the questions listed in section 3, and their results will be described in forthcoming publications.

Acknowledgements

The authors wish to thank Tomaso Poggio, Shimon Ullman, Daniel Kersten, Heinrich Buelthoff, Richard Davidson, Jim Dannemiller, Craig Berridge, Antonio Torralba, and Tabitha Spagnolo for their valuable comments on the ideas described in this paper.

References

Bar, M., and Biederman, I. (1998). Subliminal visual priming. *Psychological Science*, **9**, 464-469.

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

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

Biederman, I., and Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, **20**, 585-593.

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

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

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

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

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, **8**, 322-325.

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

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

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

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

Kanwisher, N., McDermott, J., and Chun, M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for the Perception of Faces. *Journal of Neuroscience*, **17**, 4302-4311.

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, **392**, 598-601.

Kobatake, E., and 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.

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

Newsome, W. T., and 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.

Perrett, D. I., Hietanen, J. K., Orain, M. W., and 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.

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

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

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

Sinha, P., and Poggio, T. A. (2001). High-level learning of early visual tasks. To appear in *Perceptual Learning*, Manfred Fahle (Ed.), Cambridge, MA: MIT Press.

Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 175-215.

Van Essen, D. C., and DeYoe, E. A. (1995). Concurrent processing in primate visual cortex. In M. S. Gazzaniga (Ed.), The Cognitive Neurosciences, Cambridge, MA: MIT Press.

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

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