Visual Perception and Representation of Objects and Faces

by

Javid Sadr

B.Sc. Simon Fraser University (1996)

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2003

Author ....
Department of Brain and Cognitive Sciences
September 5, 2003

Certified by ....
Pawan Sinha
Assistant Professor of Computational Neuroscience
Thesis Supervisor

Accepted by ....
Earl Miller
Picower Professor of Neuroscience
Chairman, Department Committee on Graduate Students

Visual Perception and Representation of Objects and Faces

by

Javid Sadr

Submitted to the Department of Brain and Cognitive Sciences
on September 5, 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Interpreting the results of visual object perception experiments is too often ill-posed due to the disparate and sparse choice of stimuli. To address this problem, we have developed a flexible new technique called Random Image Structure Evolution (RISE). In the simplest case, RISE involves the presentation of image sequences depicting the evolution of a coherent image from a seemingly random field, along with the reverse sequences depicting the transformation back into randomness. As it samples a subset of the space of possible stimulus images, RISE image processing strictly preserves low-level attributes such as frequency spectra and luminance, and RISE experiments are designed to provide objectively verifiable measures of the onset and offset of subjects' conscious percepts. In turn, these onset and offset measures can serve as quantitative markers for characterizing a number of intriguing perceptual phenomena. Here I describe the basic RISE paradigm and discuss experimental applications of this technique which, it is hoped, may contribute greatly to the study of key aspects of high-level vision. Building on results from psychophysical studies of perceptual onset, priming, and hysteresis, as well as findings from a magnetoencephalographic study using RISE, this thesis explores the use of RISE in characterizing the perceptual markers and neural substrates of object and face perception. In addition, this thesis examines the issue of object perception and, in particular, robust face perception, within the context of the cortical representations that may underly them, presenting and evaluating a simple, well-motivated image coding scheme based on ordinal relations.

Thesis Supervisor: Pawan Sinha
Title: Assistant Professor of Computational Neuroscience

Acknowledgments

If I have done anything of value in my life, it is due solely to the greatness of one man: Roland Fleming. (Why not?) The sun rises as the thesis sets, and a youngish man's thoughts turn to sentimental acknowledgements. Here is what comes to mind:

Say what you will, nowhere else would I have had the opportunity to receive random, brief, high-dosage exposure to Ted Adelson, who would be surprised to read this. There should be no explanation required; all I can say is I wish I'd spent more time in his presence. I have also come to greatly appreciate Emilio Bizzi's rare class, wisdom, and finesse, as well as Sue Corkin's ability to turn a lab into a sort of home, one you *can* go back to. Matt Wilson won't remember, but I did end up doing (or, rather, not-doing) as he suggested... eventually. Especially in my earlier years at MIT, Nancy Kanwisher was a great source of all things Cognitive Neuroscience and, on at least two difficult occasions, the epitome of care and propriety. I'm also grateful to have seen Mike Jordan perform his particular brand of magic twice a week for three months. There's nothing like it. It changes you. Chip Quinn and Whitman Richards deserve mad props every which way.

As a lowly research assistant aspiring to arguably-lowlier graduate studies, I had the great pleasure of getting to know Anne Treisman a little – Psychology incarnate, clothed in natural, perfect experimental design – while Danny Kahneman taught me what it takes to win a Nobel prize. Phil Johnson-Laird taught me LISP, recursion, and the lambda calculus and to this day remains in my mind as a serene embodiment of the sort of generally unattainable intellectual ideal that inspires and disheartens. Along similar lines, I'm grateful to Andy Hoffer, an uncommonly brilliant, wise, and humane man, perhaps most for simply trying to steer me in the right direction – would have worked, too, if I'd had a rudder at the time. And thanks to Bill Turnbull, I'll never listen to a conversation the same way again.

I am grateful to Patrick Cavanagh who was instantly and needlessly kind to me (and, moreover, spoke kindly *of* me), and who introduced me to the Harvard Faculty Club and assorted foreshadowing. (Also, he wrote, "But wait!") Ken Nakayama, too, with very little effort put me at great ease and has helped start the flow of some of the older, better juices. I can't thank Brad Duchaine enough for his simple, potent acts of kindness, not to mention all the DLR talk, getting the green cherries bit, and a historic brewsky run.

Certainly, I should mention that the MEG work described herein ("i before e, except after... r?") was conducted at the KIT/MIT MEG Lab under the pleasantly minimalist supervision of Alec Marantz, and I would very much like to thank Alison, Rhett, Andy, Liina, and especially Diana for making it all possible. Of course, Alison deserves more mention throughout, but I shouldn't pull another Roland.

On the administrative side, I can't say enough about the folks really running the show. Jen Vasquez, Casey, Mary Pat, Toni, Kevin, Paul, Sheila, Margo, Carla, Denise – these people deserve their own chapter. I hope it was always clear how grateful I was for each purchase order, plane ticket, reimbursement, and room-booking, not to mention the very best gossip and heads-up. Special thanks to Phyllis for our countless 3AM chats over the years. I hope retirement's treating you well.

Ken Wolff deserved a big cash bonus in general then pulled out all the stops in our last few months at Senior House – nothing would have worked out without him. And Jack Ahern got me started, but that's not even counting the curious, crafty, L33T, and lovable inhabitants: Nathan, Tanya, Cory, Zoz, Chris, Janice, Yip (hon. mention), the Finis.... The list goes on, and I know I should, sooner rather than later, write a screenplay around each of these characters.

I swear I never took for granted MIT's Athena kids, either – better than any IT department and insomniacs to boot – and reuse@mit, a ceaseless source of adrenaline, nostalgia, exercise, and much-needed home computers (not to mention more oddities than [those] at which one could shake a cricket bat). Speaking of bounty and providence, I am more than grateful to the Howard Hughes Medical Institute for the swish Pre-Doctoral Fellowship that provided me with full funding throughout my graduate career. I still remember James and I confirming each other's – that made it thrice as sweet.

I suppose that brings us to folks like Aaron, Angus, Bea, Damian, Tom, and especially Adam and Becky who were hilarious, ingenious, and just plain good to me from anear (when we were there) and afar (when I was here). These last two delinquents... an email account ain't worth beans when they don't write, and there are few other reasons to attend a conference. One should also thank one's lucky stars to have encountered Zoe, whose energy, patience, warmth, insight, and flattery are restorative; Martin, who is built like a duck's back; Barbara, even cooler in person than in research (high praise, I assure you); and Elli, who takes you by storm and, with a little help from her friends, turns VSS into Rock 'n' Roll.

Closer to home, for a while anyway, Christie was a sunshiny source of hugs, energy, laughter, and mannerisms, not to mention an unrelenting Texan host. Bill never forgot a birthday, kept a cathartic couch, made celebratory daiquiris, and taught us how to become math gods in ten pages or less. Without Serkan, the world would be a far less sexy, far less classy place, and I'm glad we didn't succeed in scaring him away. I'm not sure what I can say about Roland that wouldn't get a number of us arrested; let's just say I should like to have spent more time basking in his genius and hearty good nature. I think Elizabeth was born with the sort of grace and charm they usually have to teach royalty, and besides being brilliant and hard-working, she proved indispensable in christmastime guerilla warfare. I don't think I would have ever grown tired of Luis rubbing my beard or stubble, but I'm happy he will now see some good food and stationery. And Cindy is a saint. The least of it was letting me stay the first two weeks; more, it's her unflappable curiosity, gentle nature, classic ankles, and natural ability to fuel one's transformation into a Harsh Bastard.

Matt Tresch was why I didn't miss Jordan; Albert fostered critical thinking, let's say; Yaoda taught us how to make pork dumplings and warm sushi; Nina rejuvenated; Max deserves mention for my initiation into Matlab's inner circle, though mainly for necessitating Sayan's ultimately inconsequential eating of a lobster roll; Gina kicked me under the table; Jen was as charming and delightful as all get out, just impossible to get a hold of; Rif instantly gave me a perfect nickname; Massi somehow managed to fearlessly find his way around; Hilary gave the gift of music, the offer of late-night pizza, the keen anticipation of inbox refreshes and protracted wordplay, and even found me a place to live; Dave never did stop being a musical, Vonnegutian ray of sunshine; Sherri still hasn't explained the bread pudding and whiskey sauce; Dav harmlessly kept us on our toes and dressed up for Sayan; Itamar was scientifically badass but humble and funny; Duane (an alternate spelling for "suave"), I once heard, shared my taste for the cold.

I still think about and miss Mark, who was quiet, talented, and thoughtful and sat with me in the back row. Maybe Jaemin was the one up front, asking those engaging questions.

Gadi was god-father to us all, the voice of wisdom, experience, and dissent. I fear he is cursed with such knowledge as will not be re-discovered for decades; till then, we wait. (So what were the odds that I would chance upon and into the one thing he got so very wrong?) Neel did the heavy lifting when he didn't have to, made sure no parking ticket remained to threaten anyone else, broke more than his share of wine glasses, playfully cracked his skull *wide* open, and just sortof completed our little family here. (As a matter of fact, Neel has also reminded me to thank the ridiculous malice of the deranged, simian god who afforded me much more time than I should have had to write these acknowledgements. More affectionately and respectfully, though, I must thank Robert Frost for anticipating just about everything, for the elegant prescripts to cleanse my dear reader's palate – and for having them ready so far in advance.)

Without a doubt, I have been blessed to know the very greatest postdocs and young faculty on Earth. (Only those who have not had the pleasure could see this as hyperbole.) Hany Farid, the smoothest of the smooth, re-taught me all the signal-processing I would need for a good five years. He also taught me pool and got me the hell out of school. Federico Girosi was an inspiration, a counter-example. I've already gushed about Zoe Kourtzi. Bosco Tjan treated me like an equal from day one, and not only was he a lot of fun, he had a lot to do with getting some of this stuff right. I have come to think that Pascal Mamassian, dashing, affable, deliciously dry, will always deserve more recognition, more respect, and more rest and relaxation. Philippe Saltiel parallel-parks like the devil, and Gregg Solomon imported fresh New Year's desserts, taught me how to resolve inter-rater disputes (and Zoubin how to write acknowledgements), and re-introduced me to the world of fine writing instruments, wherein Pier taught me to press harder, fearlessly.

For the last three years, Antonio Torralba has been The Man, no doubt about it. If there's something this guy can't explain, derive, or code inside of an hour, you don't need it. Ditto for Aude Oliva. (Ditto for Sayan, too.) Guardian angels. For the first two years, Kuno Wyler-Duda was the gold standard, the best in the business, and, frankly, he always will be. Every single day, he showed me what it means to be a scientist, a mathematician, a scholar, a mentor, a gentleman, and a friend. Had I spent as much time with Petra, too, I think there would be nothing left to ask for – except for them to have stayed. Such generous, unassuming, hilarious, diligent, caring, humble, brilliant people as these deserve nothing short of everything they could ask for in life, though I know that they are also those who are content with as little or as much as they already have.

Continuing with the salts-of-the-earth: although I did a miserable job of showing it all this time, I do know that I have also been blessed with the most devoted, most affectionate friends that money can buy. As evidence, I submit first and foremost those who, abandoned and generally ignored by yours truly for about half a decade, nevertheless remained close and true, generous and demonstrative a continent's width away. Duncan didn't miss a beat and mailed a small, expensive TV to keep me company through the first long haul, and he never stopped writing, never gave up, never thought twice about driving the full length of the divide. Ken is the model of everything we could be and have if we just went for it, even if maybe sometimes he wonders if he's doing it right. He is. Angie and the boys are so sweet, so lovely and warm and bright, each time I thought I would just stay there forever, in the garage or the boat. Probably should have. Much the same goes for Steve (who taught me how things work) and Linda and the ever less-little bunch, even if in all this time I didn't drag my sorry butt there. Maybe now it'll be easier. Sean B. screamed me to the airport when it all hit the fan, cherished his shirt, and came *this* close to getting slapped after each visit to the bank. Thankfully, there was always Shari, the toothsome dental hygienist perfectly attuned to my phone voice, with whom to tease him. Even more thankfully, there is now Teri to keep him in check – *li-te-ra-lly!* Tara's dreams, waking and otherwise, were a captivating escape for us both, and Brian and Wendy, between the massive weekend breakfasts, like-clockwork birthday presents, and *truly* disgusting emails, continue to distinguish themselves as the very cream of the crop. I regularly thank the stars that Tabitha came with me that day. Boston's sad without Chad, but he made sure that I survive the first semester and establish a city-wide reputation as a controversially-costumed *bon vivant.* I was always grateful to Dr. Andrea, probably my oldest friend, for never letting me off easy, always trying to keep me in touch despite my laziness and distraction. In the same vein, I believe that Peter and I will forever stay the very best of tough-guy friends, even if neither of us ever shapes up in the email department. I thank both him and Jon for two long-ago weeks of much needed R&R at their respective homes, and I will always appreciate Cameron's simple instruction to me as he drove off – and always regret not following it. Sohrab's emails were fun, enthusiastic, little-brotherly treats, though I'm still trying to remember if I saw him more than once this whole time (and where the hell's that book, dude?), and Richard B. and Peter C. forgave my year-long silences with prompt and jovial replies. I think I should have stayed with you guys, but we would definitely have had to migrate to linux. Through Jackie, my lovely surro-sis, I vicariously traveled the world, met the world's strongest men, and held on to my biomedical roots and SFU ties. I hope you don't forget me when your healing hands are fully trained and certified. Trina flipped the coin, then stayed (from afar) to weather the choice. Against all odds, she brightened each Christmas I let her lay her hands on, never asking for anything in return, forgiving the long silences, and always just so very beautifully kind. I was always matter-of-factly invited to Morgan's tryptophan resistance training dinners, whether or not I was on the correct coast, and the first time I made it, she met me so warmly... it was nothing short of a spiritual experience. And Christine... I honestly can't even begin to do her justice.

In Carolina, Sam fed me beef that broke my heart and biscuits that saved my soul and, whether riding around in his pick-'em-up truck or lounging on his monumental patio, made me feel more at home than at home. I could cry, thinking of how I met his and Stephanie's hospitality, fit for a king, with an antisocial retreat into last-minute paper-writing. Well, you guys, check out Chapter 2 and decide if it was worth it – and, seeing that it probably wasn't, please invite us back. They, along with Ron, truly the gentleman's gentleman, and Maria, now legally bound to keep him out of trouble, will be forever in my heart, and I, forever in their debt, for all they did in Tabitha's second and third years there. And for one year, when it seemed like it was mainly just Ron, I thanked heaven for him every day. I can never repay them, and it all does have a lot to do with this thesis – and with hers.

On to my beloved family, where I should like to start with the freshest. Dr. John Spagnolo is probably the first professional scholar I got to observe at close range, and that, as much as anything else at the time, had a lot to do with my not stopping with a B.Sc. Oh that one should find professors of this calibre so few and far between, or what's a Heaven for? My esteemed MIL, Penny, readily divined the way to my heart, fabricated dirty rumors of me and a mysterious blonde, withstood the natives, and is just plain cool. One hopes that her memoirs are finally in the works. Becky loves me better than any sister I could have asked for – and protective? I will not name those who owe their lives to my holding her back. (Ditto for Sayan. Again, not hyperbole.) I've lost track of everything she's done for us, and I know I'll never be able to pay her back... but here's a start: "Screw you guys, I'm going to Harvard!" So, then, Gonzalo fits here, too, and not only for always believing in my Kung Fu/foo; among many other things, we have him to thank for the most beautiful picture(s) to follow.

My father probably had the most to do with my going this far. This is the sort of fellow who has (effortlessly) memorized more and better prose and poetry than is stocked in your average city library, who has written more than will ever be discovered or rightly attributed – a natural, unadulterated intellect who deserved and would have had a Ph.D. himself, many times over, were it not for his uncommon sense of duty and sacrifice. This one's for him, I'd be inclined to say; more to the point, however, I do suggest he now stick a fork in me. My mother, who gave me at least half of what passes for my brains (and face, actually), is the sort that all but swims through an M.D., whose mathematical prowess spawns chilling urban legends, whose blind white-on-rice love for her son weathers such perennial disappearances as those involved in the making of a Ph.D. (No animals were hurt, but I doubt it was easy on the folks.) She will see the gobbledygook for what it is but, I hope, will still get a hell of a lot of mileage out of it. Jalal raised the bar improbably high, most likely as a dare, though maybe he didn't know it at the time, and I guess we can say it worked. You want a coach for a high-school math contest, someone to show you the ropes in university, or maybe just a high-speed ride to your chemistry final two hours late? This is the guy. My number one border-crossing commiserant – and he cleans up pretty damn good for weddings, too. (Ditto for Sayan.) Omid, who knew better than to fall into the same trap as the first two, or at least to not keep digging, continues to inspire and harangue as he lives the life. I hope he calls before I have the heart-stopping surprise of seeing him at a Cannes press junket or on this week's MTV Cribs... or, much more likely, Punk'd. I know it won't be long, unless something even bigger and badder is in progress. Who else? A little further out, Marvin taught me some smooth moves I dare not ever use, and Nabil was my eight-year-old guide to western culture, a sort of pint-sized Iggy Pop.... And one sunny afternoon in the front yard, my aunt Vedad told me that the stars are always shining up there, that you just can't see them when it's bright.

James and Sayan – their company these last not-so-few years has been, let us say with understatement and euphemism, formative. I've learned that I was born to drive (and want the world at my 6), that sometimes you leave and sometimes you fight, and to just drink. We made our own fun – what else was there? – and somehow we never got caught. I hope they will have the opportunity now to get to know me, again.

–

Finally, and most of all, I must thank, and forever continue to thank, my wife, Tabitha, who bore the weight and suffered the worst, who lost even more sleep and innocence than I. If any of these sentences conform in any way to the accepted rules of English grammar, you know who to thank. (Little known fact: it is her perfect and complete grasp of *French* grammar that serves as the basis of her most arcane but technically correct formulations, but you've guessed wrong if you think that means she's French.) She was always my most conscientious and insightful test pilot, my most patient and observant vocal coach, my smartest (sometimes smartingest) critic. I know that words only diminish the truth of her, but I can at least state my wish to someday deserve her, to someday make up for all this, to someday make good on all the vows and promises. She was the brains of the operation and, after a time, even the brawn.

–

I had thought I might dedicate this thesis, whatever its worth, to my father or to my wife, the two who paid the most for it, but he could have written ten better of his own (he may yet) and she will have a proper one in no time. I think instead that it should be for Kaz. For Kaz because he tried so hard, because there wasn't enough time, because he was a joy, because he still keeps me company,... because *he* would have thought it's cool.

Contents

1 Introduction **16**

1.1 Background and Motivations . . . 16

1.2 Object Perception and Random Image Structure Evolution . . . 16

1.3 Exploring the Neural Correlates of Object and Face Perception using Magnetoencephalography and Random Image Structure Evolution . . . 18

1.4 The Fidelity of Local Ordinal Encoding . . . 20

1.5 Conclusions . . . 20

2 Object Perception and Random Image Structure Evolution **22**

2.1 Introduction . . . 22

2.2 The RISE Paradigm . . . 25

2.2.1 Stimuls Image Processing . . . 25

2.2.2 Basic Experimental Paradigm . . . 35

2.2.3 Experiment A: Perceptual Onset, Offset, and Hysteresis . . . 38

2.2.3.1 Aim . . . 38

2.2.3.2 Methods . . . 38

2.2.3.3 Results . . . 39

2.3 Applications of the RISE Paradigm . . . 40

2.3.1 Characterizing the Neural Substrates of Object Perception . . . 40

2.3.2 Quantitative Assessment of Priming . . . 45

2.3.2.1 Experiment B: Priming Perceptual Onset . . . 46

2.3.2.1.1 Aim . . . 46

2.3.2.1.2 Methods . . . 46

2.3.2.1.3 Results . . . 47

2.3.3 Developmental and Clinical Studies 49

2.3.3.1 Normal and Abnormal Perceptual Learning and Development 49

2.3.3.2 Visual Agnosias and Prosopagnosia 50

2.3.4 Studying Top-Down Influences on Early Visual Areas 50

2.4 Discussion 51

3 Exploring the Neural Correlates of Object and Face Perception using Magnetoencephalography and Random Image Structure Evolution **57**

3.1 Introduction 57

3.2 Methods 61

3.3 Results 64

3.4 Discussion 69

4 The Fidelity of Local Ordinal Encoding **74**

4.1 Introduction 74

4.2 Information Content of Local Ordinal Encoding 77

4.3 Discussion 79

5 Conclusion **84**

5.1 Contributions 84

5.2 Future Directions 85

List of Figures

1-1 A subset of a RISE sequence in which an image with random phase is gradually transformed into the infamous Lena. 17

1-2 MEG data from a typical subject, illustrating both the modulation of the M100 and M170 to the onset of face perception (appearance of these components at level 3 when perceptual onset was reported) and the enhancement effect of hysteresis on the neural activity on subsequent viewing of the same image (level 5). These data taken from three occipitotemporal sensors independently determined to be selective to face stimuli. Similar results are seen for object stimuli, but more commonly with a suppressive effect of repetition. 19

1-3 The image on the right is a reconstruction of the source image after it was encoded purely as a set of ordinal (i.e., greater than, less than) relations between neighboring pixels. The reconstruction was performed using a simple, biologically plausible, relaxation-based technique. 20

2-1 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. 24

2-2 A sample RISE sequence generated by pair-wise 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. 26

2-3 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. 30

2-4 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 -π to π radians). Bottom row: output images from inverse Fourier transform given original amplitude spectrum and modified phase spectrum randomized in the (d) 1 Hz - 16 Hz, (e) 17 Hz - 32 Hz, and (f) 33 Hz - 48 Hz frequency bands. 31

2-5 Ten RISE sequences were generated, each based on one of ten 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. 35

2-6 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. 37

2-7 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. 41

2-8 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. 48

3-1 A subset of a RISE sequence in which an image with random phase is transformed into an easily recognized photograph of a face. While the technique progressively transforms the spatial structure of the source image, a number of important low-level characteristics of the image, such as its overall spatial frequency spectrum, luminance, and contrast, are controlled throughout. 59

3-2 Sample RISE sequence depicting four frames at equally spaced levels of degradation. Each four-frame series was converted to a seven-frame sequence for presentation in the experiment and, further, interleaved with another randomly selected seven-frame sequence (see text). 62

3-3 Time-course of each presented frame: 150 - 250ms fixation period, 600ms image presentation, 800ms post-presentation response period (note that subjects could also respond during image presentation), 200ms interstimulus interval. Subjects were asked to fixate and not blink or move their eyes during the fixation and image presentation periods. 63

3-4 Grand means and standard errors of peak amplitudes for each perceptually-based trial category, face stimuli. The results consistently show preonsets < preonset-complements, preonsets < onsets, and onsets < level 4 (best images). M170 measures also indicate a neural enhancement effect for face stimuli (e.g., compare onsets and onset-complements, right hemisphere). (a) M100 troughs (left hemisphere, dark gray, absolute value shown) and peaks (right hemisphere, light gray). (b) M170 peaks (left hemisphere, dark gray) and troughs (right hemisphere, light gray, absolute value shown). See text for statistical tests between conditions. 65

3-5 Grand means and standard errors of peak amplitudes for each perceptually-based trial category, object stimuli. The results consistently show preonsets < preonset-complements, preonsets < onsets, and onsets < level 4 (best images). M170 measures also indicate a neural suppression effect for object stimuli (e.g., compare onsets and onset-complements). (a) M100 troughs (left hemisphere, dark gray, absolute value shown) and peaks (right hemisphere, light gray). (b) M170 peaks (left hemisphere, dark gray) and troughs (right hemisphere, light gray, absolute value shown). See text for statistical tests between conditions. 66

3-6 Grand means and standard errors of the latencies of peak amplitudes for each perceptually-based trial category, face stimuli. Please see text for results of statistical comparisons between conditions. 67

3-7 Grand means and standard errors of the latencies of peak amplitudes for each perceptually-based trial category, object stimuli. While suggestive of an earlier M100 latency for the level 4 (best) images, no comparisons here yielded significant differences (except that the M170 tends to peak earlier in the left hemisphere as compared to the right). 68

4-1 (a) A schematic contrast response curve for a primary visual cortex neuron. The response of the neuron saturates at low contrast values. (b) An idealization of (a). This unit can be thought of as an ordinal comparator, providing information only about contrast polarity but not its magnitude. 75

4-2 The challenge for a representation scheme: to construct stable descriptions of objects despite radical changes in appearance. 75

4-3 The absolute brightnesses and their relative magnitudes change under different lighting conditions but several pair-wise ordinal relationships stay invariant. 76

4-4 Ordinal relationships between an image region p_a and its neighbors. 77

4-5 Reconstruction results from the regularization approach. (a) Original images. (b) Reconstructed images. (c) Absolute difference between original and reconstruction. (d) Histogram of absolute difference. 79

4-6 Reconstruction results from the relaxation approach. 80

4-7 Five reconstructions, shown here to demonstrate the robustness of local ordinal encoding to missing inputs. From left to right: reconstructions based on 100%, 80%, 60%, 40%, and 20% of the full set of immediate neighbor relations. 81

4-8 A small collection of ordinal relations (a), though insufficient for high fidelity reconstruction, is very effective for pattern classification despite significant appearance variations. (b) Results of using a local ordinal relationship based template to detect face patterns. The program places white dots at the centers of patches classified as faces. (From Thoresz and Sinha, in preparation.) 82

We dance 'round in a ring and suppose,
but the Secret sits in the middle and knows.

- Robert Frost -

Chapter 1

Introduction

1.1 Background and Motivations

Much of the work described in this thesis is grounded on the notion that psychophysical and neuroimaging/eletrophysiological research of high-level vision can be greatly served by a simple but well-principled approach to stimulus generation/selection, behavioral paradigm, and subsequent data analytic methodology. Each research chapter describes, in turn, techniques and experiments conducted in the psychophysical exploration of object perception, the magnetoencephalographic study of perceptually relevant early visual responses to face and object stimuli, and the development of a simple but biologically motivated image coding scheme that may underlie the neural representations supporting these perceptual processes. Each chapter discusses a different sub-field of inquiry, and does so with reference to a good deal of research directly relevant in each domain. Therefore, the current section does not recapitulate an omnibus review of all key issues and literature. Rather, what immediately follows covers a concise overview of the main work covered in this thesis, and each of the chapters individually provides the reader with the requisite grounding in the background, motivations, problems, and research specifically related to each subtopic.

1.2 Object Perception and Random Image Structure Evolution

The study of object perception has its basis in the characterization of behavioral and/or neural responses to images of different objects. However, such images can be thought of as isolated points drawn from a high-dimensional image space, and response measures corresponding to these points do not provide information regarding variation in response associated with systematic variation in image attributes around these points. In the case of face perception, for example, a set of neurons or brain regions which may 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 "near" face images in this image space but that are not in fact perceived as faces or even face-like. Our aim is to develop a flexible new approach to the study of object perception, one which is based on the ability to systematically probe responses with sets of images that lie along continuous paths in image space.

In its most basic form, our technique (called RISE, for Random Image Structure Evolution) involves the evolution of a random field into an image of an object, as well as the reverse sequence depicting the progressive transformation of the object's image back into randomness. A subset of a sample RISE sequence is shown in figure 1-1. Throughout the RISE sequence, our image processing strategy preserves low-level image attributes such as the frequency spectrum and luminance, and it also allows for other visual effects, such as degradation of local features and morphing between images of different objects. Experiments based on the RISE paradigm can be used to address a number of key issues in object perception.

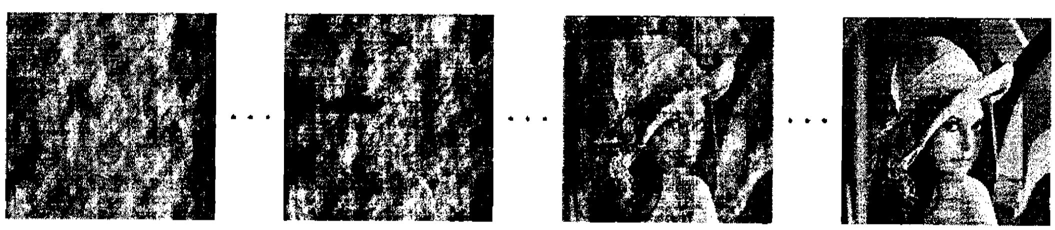


Figure 1-1: A subset of a RISE sequence in which an image with random phase is gradually transformed into the infamous Lena.

In the simplest RISE experiment, we can objectively quantify the onset and offset of object recognition as subjects view images evolving from, then dissolving into, randomness. Shifts in onset and offset across different conditions or populations, then, can serve as reliable measures of various perceptual phenomena. In our experiments thus far, we have used onset shifts as new indices of object priming, obviating the need for precise reaction-time measurements or tachistoscopic presentations. We have also been able to quantify a marked hysteresis effect: the offset of recognition occurs at a level of image degradation far worse than which supports the onset of recognition.

Beyond simply validating the findings of this technique against existing literature, we show how RISE may be used to unambiguously uncover subtle perceptual effects that may be relatively inaccessible by other experimental approaches (for example, due to the relatively large motor response variability inherent in reaction time tasks). We also explore how the objective and sensitive markers from RISE may be translated into useful measures of perceptual development and learning, as well as sensitive tests of visual deficits such as visual agnosia. Finally, by establishing links between behavioral reports and neural responses, and as discussed in more detail in the following chapter, we show how RISE may be used to investigate the neural correlates of object recognition, as well as those of priming/repetition and perceptual hysteresis. (This chapter is a preprint of a research paper that is to appear in an upcoming special issue of the journal Cognitive Science.)

1.3 Exploring the Neural Correlates of Object and Face Perception using Magnetoencephalography and Random Image Structure Evolution

There is currently great interest in the mechanisms underlying object and face perception, and it is not uncommon to approach their study through the characterization of functional and structural neural correlates that appear differentially responsive to certain object classes. However, it may be difficult to dissociate neural activity driven by low-level stimulus features from that driven by higher-level object percepts. We explore an alternative approach in which changes in perceptual and neural activity are measured as visual stimuli undergo systematic transformations (e.g., as recognizable objects evolve from and then dissolve into randomness) while important low-level image properties are preserved.

Subjects' percepts often arise sharply in such circumstances and thereafter exhibit a marked hysteresis, and these properties, along with the ability to objectively verify subjects' reported percepts, offer distinct advantages when this technique is integrated with neuroimaging or electrophysiology (here, magnetoencephalography) to isolate perceptually relevant neural activity. Such activity should demonstrate a similar time course, one with a relatively sharp onset as well as a delayed offset consistent with the perceptual hysteresis. With this approach we do indeed observe the modulation of early visual responses (M100/P120 and M170/N200 components) coincident with explicit perception of objects and faces (an illustration of one subject's responses appears in figure 1-2), as well as an intriguing response enhancement/suppression dichotomy between these two stimulus classes. Similar experiments may help further our understanding of the neural mechanisms underlying a number of important perceptual phenomena.

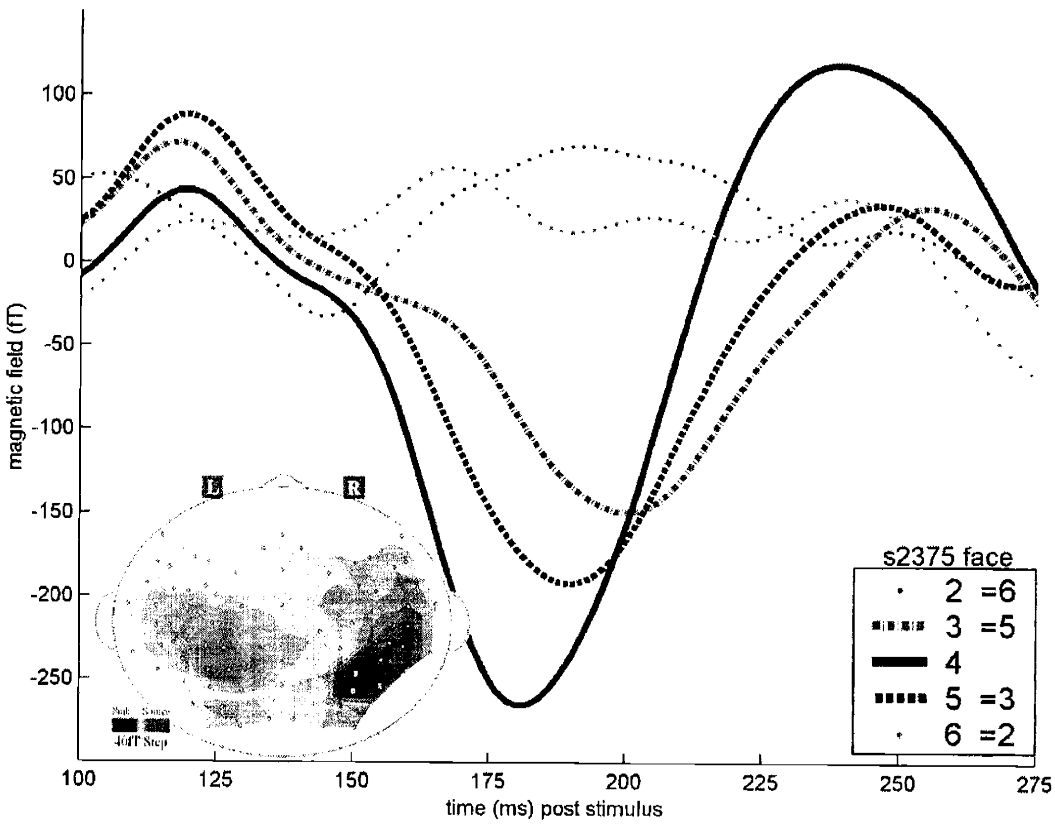


Figure 1-2: MEG data from a typical subject, illustrating both the modulation of the M100 and M170 to the onset of face perception (appearance of these components at level 3 when perceptual onset was reported) and the enhancement effect of hysteresis on the neural activity on subsequent viewing of the same image (level 5). These data taken from three occipitotemporal sensors independently determined to be selective to face stimuli. Similar results are seen for object stimuli, but more commonly with a suppressive effect of repetition.

1.4 The Fidelity of Local Ordinal Encoding

A key question in neuroscience is how to encode sensory stimuli such as images and sounds. Motivated by studies of response properties of neurons in the early cortical areas, we propose an encoding scheme that dispenses with absolute measures of signal intensity or contrast, using instead only local ordinal measures. In this scheme, a signal's structure is represented by a set of equalities and inequalities across adjacent regions. A sample reconstruction is depicted in figure 1-3.

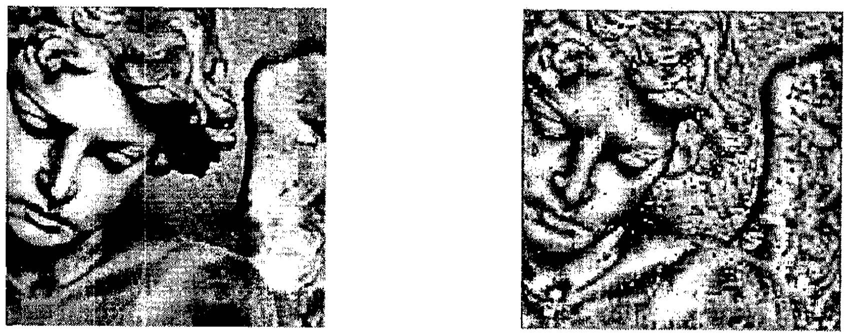


Figure 1-3: The image on the right is a reconstruction of the source image after it was encoded purely as a set of ordinal (i.e., greater than, less than) relations between neighboring pixels. The reconstruction was performed using a simple, biologically plausible, relaxation-based technique.

Here, we focus on characterizing the fidelity of such representations. We develop a regularization approach for image reconstruction from ordinal measures and thereby demonstrate that the ordinal representation scheme can faithfully encode signal structure. We also present a biologically plausible implementation of this computation that uses only local update rules. The results highlight the robustness and generalization ability of local ordinal encodings for the task of pattern classification. (This chapter is a reprint of a research paper included in the proceedings of the 2001 Neural Information Processing Systems conference.)

1.5 Conclusions

In the final chapter, we provide a concise overview of some key contributions of this thesis, both in terms of new methodological developments and of important psychophysical and electrophysiological findings. In closing, we discuss a number of important and intriguing directions for future research.

Let chaos storm!
Let cloud shapes swarm!
I wait for form.

- Robert Frost -

Chapter 2

Object Perception and Random Image Structure Evolution

2.1 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 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., Puce, Allison, Gore & McCarthy, 1995; Kohler, Kapur, Moscovitch, Winocur & Houle, 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 $\mathbf{x}$ if too few data points exist, if these points are distributed too unevenly along the $\mathbf{x}$ axis, or, worse, if there is no formal metric by which to represent the variations in $\mathbf{x}$ 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 figure 2-1, images can be thought of as points lying in a high-dimensional space, where each dimension corresponds to a way in which images may vary. (For instance, a 100x100 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 images 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 relating 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 reflect 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. This is the key motivation underlying the RISE paradigm.

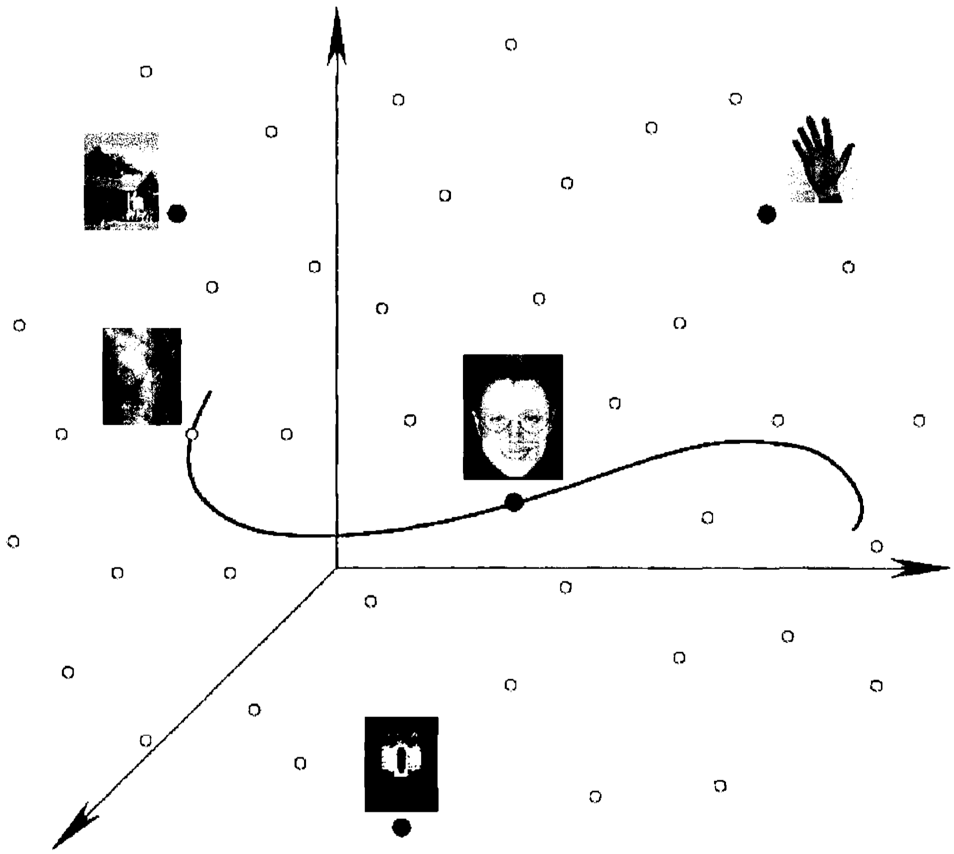


Figure 2-1: Conceptually, a given image can be said to correspond to a point lying in a high-dimensional "image space," as shown here schematically. Specific images of objects and scenes, of the kind used in most recognition studies constitute a very sparse subset of points in this space (filled circles). 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.

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.

2.2 The RISE Paradigm

2.2.1 Stimuls 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., Snodgrass & Corwin, 1988; Dolan et al., 1997; James, Humphrey, Gati, Menon & Goodale, 2000; 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 figure 2-2 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.,

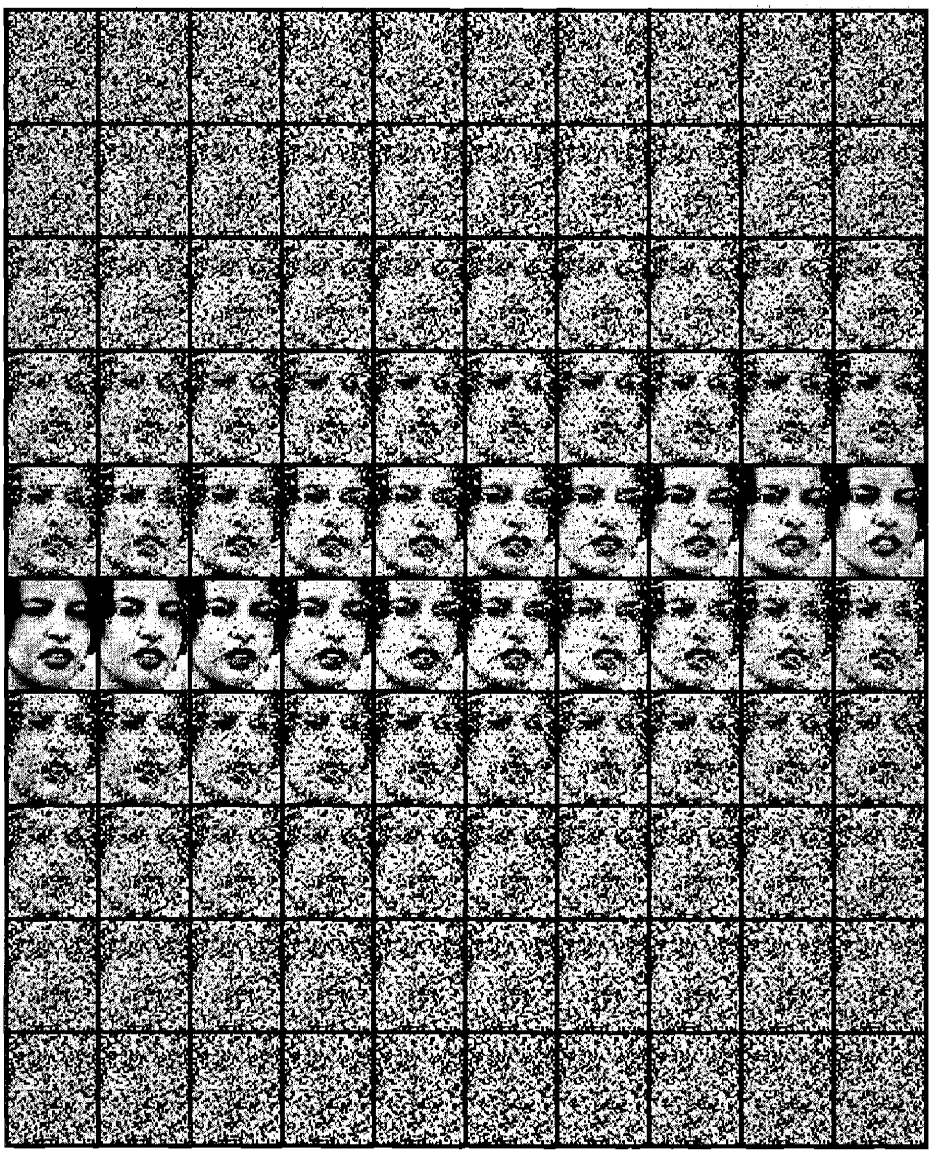


Figure 2-2: A sample RISE sequence generated by pair-wise exchanges of image regions (here single pixels). A simple presentation of these images would proceed in raster order (i.e., from left to right and top to bottom). The source image appears in the first column of the sixth row.

Grill-Spector, Kushnir, Hendler, Edelman, Itzchak & Malach, 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 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 eight to sixteen 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 but all 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 versus flexible coarse-to-fine strategies of spatial scale usage) can be found in Morrison and 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., Legge, Pelli, Rubin & Schleske, 1985; Gold et al., 1999; 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 and makes 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 this 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 low-level attributes of the resulting image will 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 matrix combined with the amplitude 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 2-3 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. Figure 4 shows how such a decomposition might look for a two-dimensional signal such as a 100x100 pixel image (figure 2-4.a). 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 (figures 2-4.b and 2-4.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

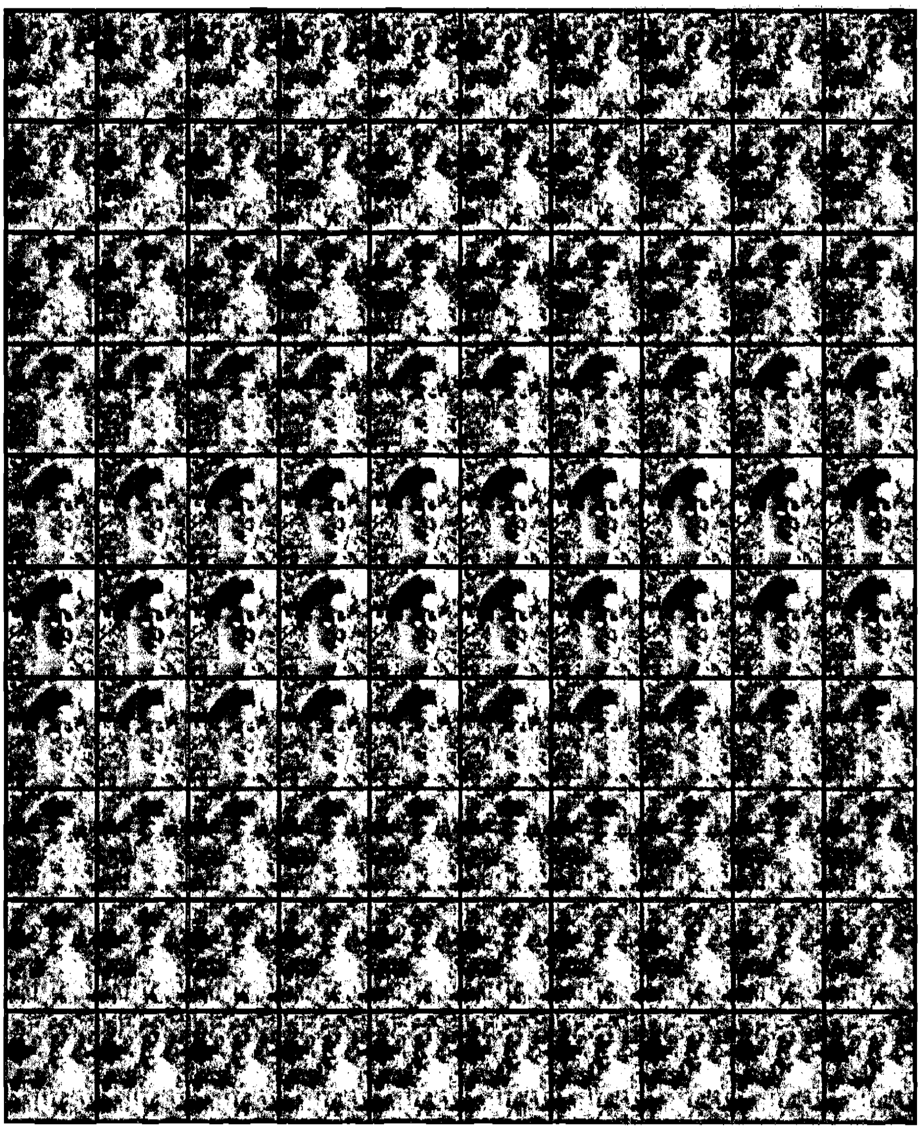


Figure 2-3: A sample RISE sequence generated by progressive degradation of the phase matrix of the source image. Although the technique progressively disrupts the spatial structure of the image, important low-level image properties of the original image, such as the spatial frequency spectrum and overall luminance, are perfectly preserved.

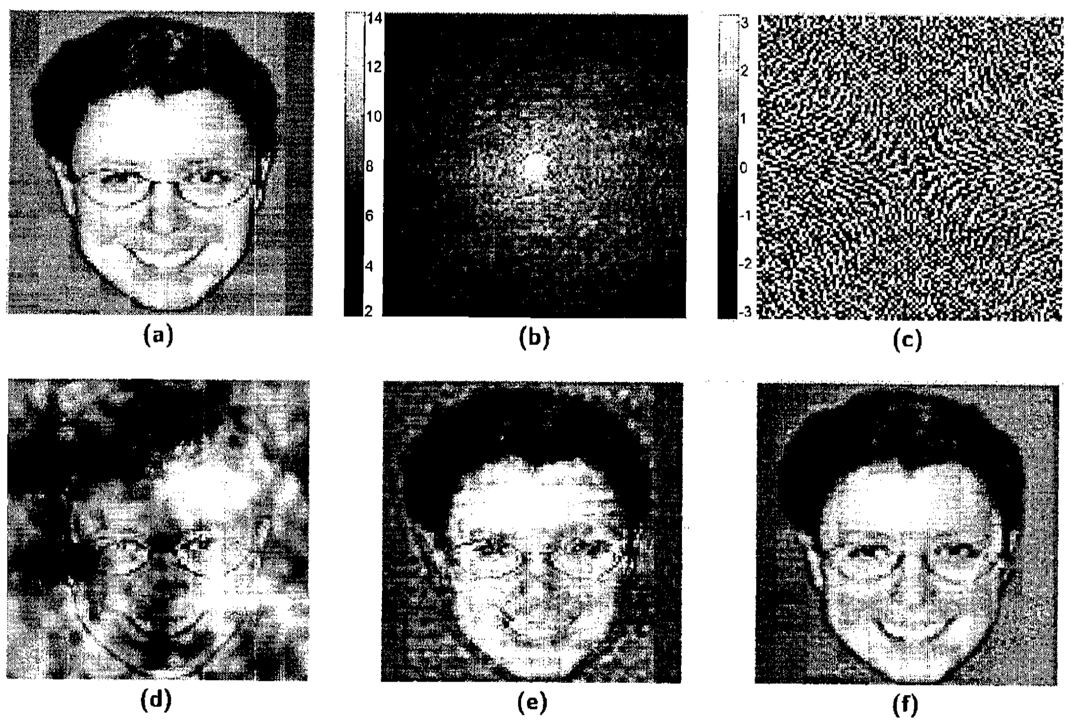


Figure 2-4: Simple illustration of Fourier analysis and phase randomization. Top row: (a) original image, (b) log of amplitude spectrum, and (c) phase spectrum (values range from -π to π radians). Bottom row: output images from inverse Fourier transform given original amplitude spectrum and modified phase spectrum randomized in the (d) 1 Hz - 16 Hz, (e) 17 Hz - 32 Hz, and (f) 33 Hz - 48 Hz frequency bands.

Hz 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 usually 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. Figures 2-4.d, 2-4.e, and 2-4.f depict the output of the inverse Fourier transform following the randomization of phase values in specific frequency bands (1 Hz - 16 Hz, 17 Hz - 32 Hz, and 33 Hz - 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. 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 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 + 2π 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 figures 2-4.b and 2-4.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 the 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 (figure 2-5). 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 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

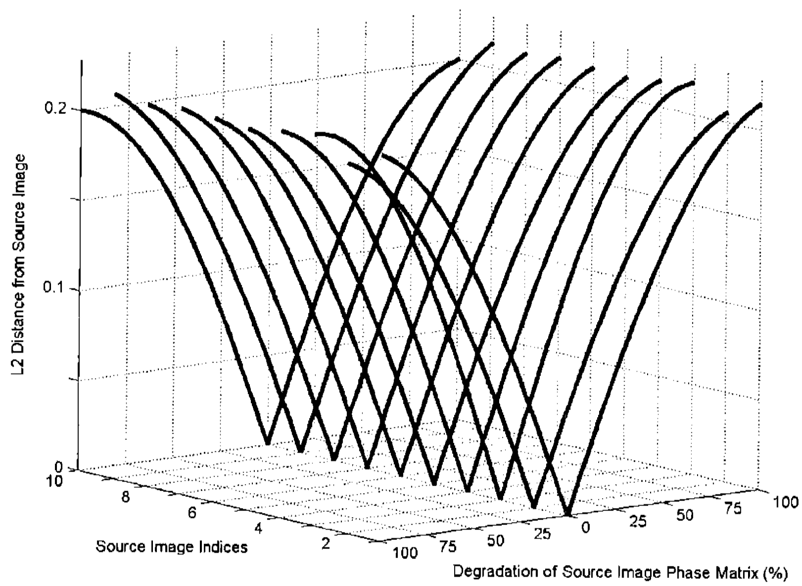


Figure 2-5: Ten RISE sequences were generated, each based on one of ten source images common objects. Here are plotted the L2 distances (i.e., sum of squared differences) between each of the original source images and the images in its associated RISE sequence. It can be seen that RISE image processing 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.

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.

2.2.2 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 two important aspects of an observer's percepts. The first of these can 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 (figure 2-6). 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 can not 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 per-

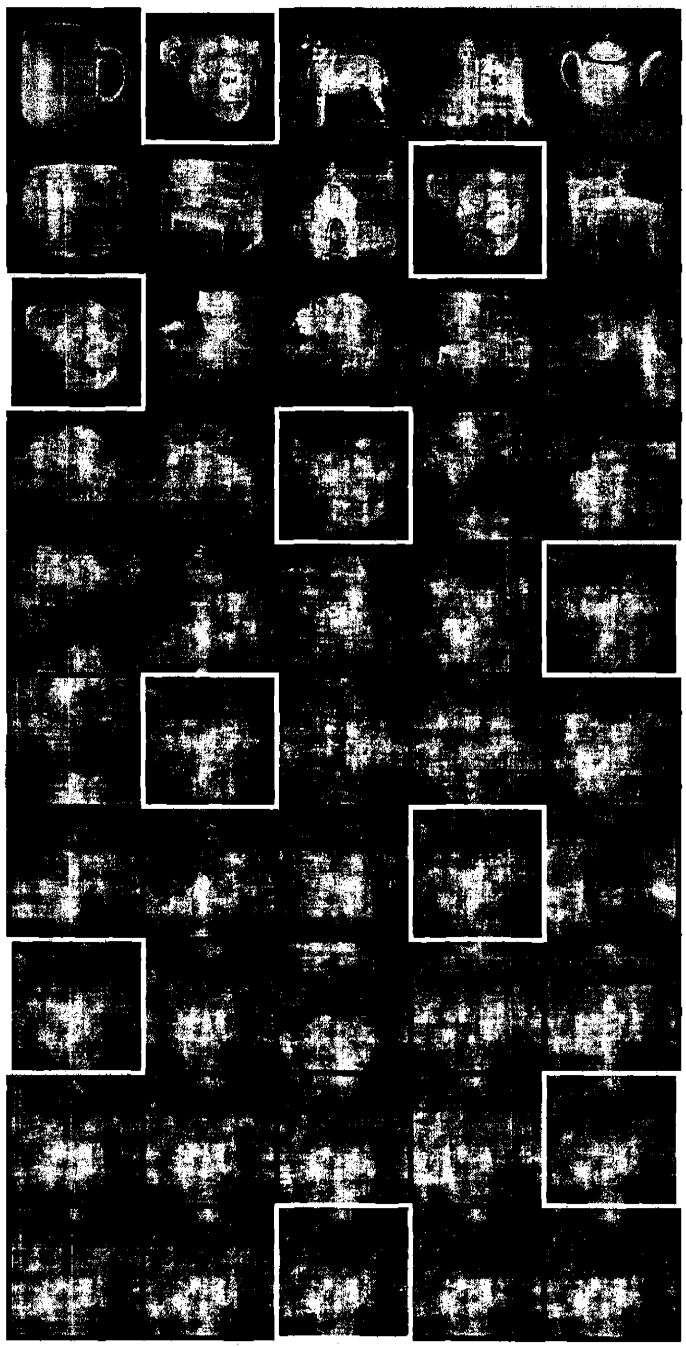


Figure 2-6: A sample "mixed" RISE offset subsequence incorporating distractors. Here we have indicated the target object by framing it in white. Throughout the sequence, the images have precisely the same luminance and frequency amplitude spectrum (Fourier magnitude). Here, they also gradually converge to a common random phase spectrum.

ceptual 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.

2.2.3 Experiment A: Perceptual Onset, Offset, and Hysteresis

2.2.3.1 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.

2.2.3.2 Methods

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.

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 ten distractors. These sequences were generated from 255x255 pixel grayscale images of easily identified objects (see figure 2-7 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 fifteen 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.

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 candela/m^2; maximum: $\sim$90 candela/m^2) in a dimly lit room ($\sim$5-10 candela/m^2) and were viewed binocularly. Each image in these sequences was presented for 750ms and subtended approximately 8 degrees of viewing angle. At the end of each 750ms 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 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.

2.2.3.3 Results

Figure 2-7 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, $F(1, 30) = 101.4$, $p \ll 0.001$. This result and its implications are discussed further below.

2.3 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.

2.3.1 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; Zeki, Watson, Lueck, Friston, Kemard & Frackowiak, 1991; Van Essen & DeYoe, 1995).

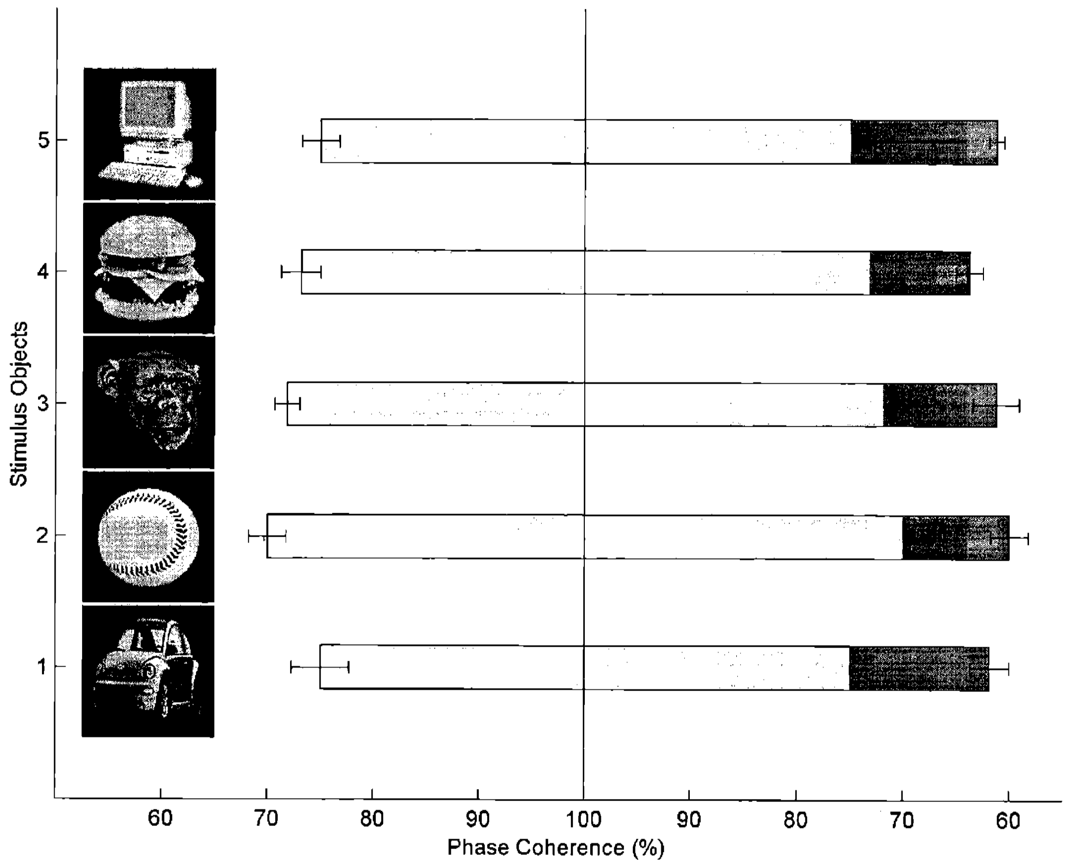


Figure 2-7: Onset and offset of object perception in RISE sequences for five objects, represented by the left and right edges of each bar, respectively. Observers exhibit a marked perceptual hysteresis during the offset subsequence – the transition from light to dark gray in each bar serves to indicate the reflection of the onset point about the vertical axis. Data are averaged across four observers.

However, in terms of the neural substrates of more complex perceptual faculties, though much progress has been made over the past several years (Perrett et al., 1992; Martin, Wiggs, Ungerleider & Haxby, 1996; 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, Harel, Hendler, Ben-Bashat, Zohary & Malach, 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 versus 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 (100ms) 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, 2003a, 2003b), 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. These experiments are the focus of an upcoming paper (Sadr & Sinha, 2003b).

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).

2.3.2 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 (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.

2.3.2.1 Experiment B: Priming Perceptual Onset

2.3.2.1.1 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.

2.3.2.1.2 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.

2.3.2.1.3 Results

Figure 2-8 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, $F(1, 30) = 34.34$, $p \ll 0.001$. 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, $F(1,30) = 26.2$, $p \ll 0.001$ – 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

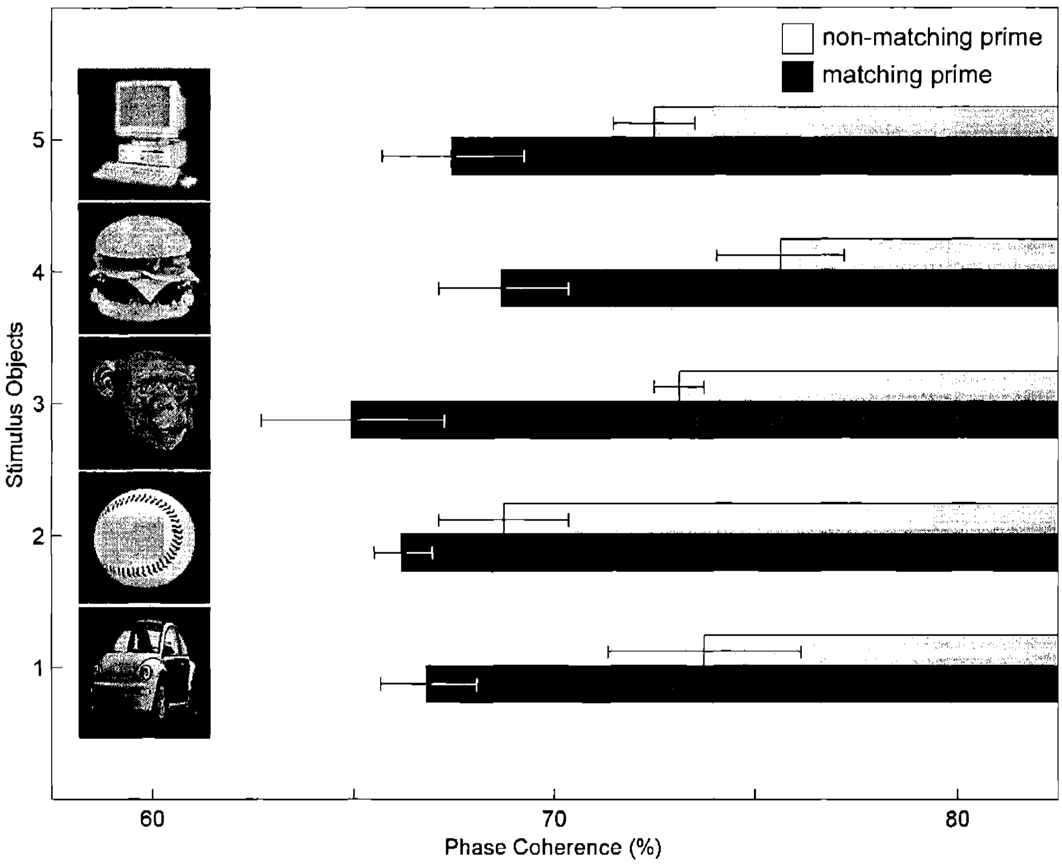


Figure 2-8: Comparing onset of object perception with matching versus unrelated verbal priming induced by a concurrent word memory task. Black bars correspond to mixed RISE presentations verbally primed by a word matching the target objects' name, while light gray bars correspond to the same RISE sequences 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.

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 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.

2.3.3 Developmental and Clinical Studies

2.3.3.1 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 (Brooks & Goldstein, 1963; Diamond & Carey, 1986; Bartlett & Searcy, 1993). 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 and a reduced sensitivity to sadness cues (Pollak & Sinha, 2002).

2.3.3.2 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 (Farah, 1990; Damasio, Damasio & van Hoessen, 1982), 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.

2.3.4 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 (Sinha & Poggio, 1996, 2002; Jones, Sinha, Poggio & Vetter, 1997). 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.

2.4 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 has been 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; Murray, Bennett & Sekuler, 2002; Mangini & Biederman, 2001). 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., Pomplun, Ritter & Velichkovsky, 1996; Gosselin & Schyns, 2001). 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 3D 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 3D 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, 3D 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 versus 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.

I often see flowers from a passing car
that are gone before I can tell what they are.

I want to get out of the train and go back
to see what they were beside the track.

I name all the flowers I am sure they weren't;
not fireweed loving where woods have burnt –

not bluebells gracing a tunnel mouth –
not lupine living on sand and drouth.

Was something brushed across my mind
that no one on earth will ever find?

Heaven gives its glimpses only to those
not in position to look too close.

- Robert Frost -

Chapter 3

Exploring the Neural Correlates of Object and Face Perception using Magnetoencephalography and Random Image Structure Evolution

3.1 Introduction

An interesting and important field of inquiry in neuroscience is the exploration of the mechanisms by which the human visual system performs object and face perception. In this domain, there has been recently something of an explosion of research, especially work focused on face perception. Combined with the advent and proliferation of both more advanced neuroimaging techniques, e.g., functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG), and development of intriguing new behavioral paradigms, this work has added greatly to the existing electrophysiological and psychophysical literature on this subject, especially in terms of the time-course and localization of the neural activity associated with high-level percepts. Nevertheless, numerous gaps remain in the understanding of key issues in this domain, and, perhaps more importantly, conceptual and methodological shortcomings of certain approaches continue to fuel controversy and diminish the impact of some studies. Our aim in this paper is to directly explore the neural correlates of object and face perception, along with their modulation by perceptual learning and hysteresis, and to do so using a principled and integrated approach to stimulus generation, experimental protocol, and neural recording.

Undeniably, neuroimaging research of high-level vision perception has been developing a particular focus on face perception in recent years, and fMRI, EEG, and MEG work in this domain has made great strides in characterizing the neural correlates of this most remarkable visual faculty. Perhaps the most attended work has been that which has proposed a face-selective cortical area in the fusiform gyrus (e.g., Kanwisher, McDermott & Chunn, 1997) and, moreover, that which has described the very early time-course of neural activity best associated with face processing (e.g., Linkenkaer-Hansen et al., 1998; Allison et al., 1999; Puce et al., 1995; Seeck et al., 1997; and others). The latter work, typically performed using MEG and/or surface EEG recordings, has demonstrated what appears to be a face-selective occipitotemporal signal component at approximately 160 - 190ms after stimulus onset (often referred to as the N170 or M170; Linkenkaer-Hansen et al., 1998; Allison et al.,1999; Liu et al. 2000; and others), along with an still-earlier component at approximately 100 - 120ms (referred to as the P120 or M100; Linkenkaer-Hansen et al., 1998; Halgren et al., 2000; Liu, Harris & Kanwisher, 2003; and others).

The face-selectivity of these two signal components – that is, their significantly greater amplitude in response to images of faces as opposed to other objects – has been a boon to face perception research, but it has, in a fashion, at times cast object perception into the role of a control task. That is, many fMRI and MEG researchers routinely use the neural response to object images solely as a subtractive baseline by which to assess the relative amplitude, and subsequent modulation, of the face-specific response. Further, as discussed in related work (e.g., Sadr & Sinha, 2001, 2003, in press), such comparisons of face versus non-face images typically embody a theoretical limitation in their lack of a continuous independent variable (i.e., the stimuli represent discrete categories against which it would be difficult to plot a strictly-defined response function). Along these lines, we have recently developed a technique whereby any image of an object, face, scene, etc., can be systematically transformed into an ordered set of increasingly (or decreasingly) degraded stimuli while conserving a number of important low-level image properties (e.g., luminance, spatial frequency spectrum, contrast; see figure 3-1 for a sample sequence). This technique also incorporates a simple behavioral paradigm by which the perceptual and/or neural correlates (e.g., perceptual onset/offset, appearance of M100/M170 component, etc.) of a subject's viewing of the image progression can be assessed, independent of the use of other object/face comparison images.

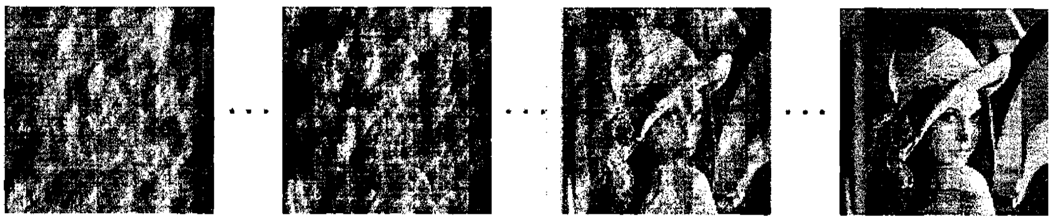


Figure 3-1: A subset of a RISE sequence in which an image with random phase is transformed into an easily recognized photograph of a face. While the technique progressively transforms the spatial structure of the source image, a number of important low-level characteristics of the image, such as its overall spatial frequency spectrum, luminance, and contrast, are controlled throughout.

A key issue related to this approach and, more importantly, to object and face perception in general, is the influence of visual experience on perception and neural activity. For example, if within a series of presented images, an item is repeated, how will responses to the second presentation of the item differ from the first? Using fMRI and a randomized, serial presentation of grayscale drawings of objects, Buckner et al. (1998) have shown that extrastriate, inferotemporal, and left dorsal prefrontal cortical activity correlated with object perception is reduced for repeated stimuli. Rugg and Doyle (1994), employing a repetition-detection task, demonstrated a similarly enhanced EEG positivity after approximately 200 - 250ms for repeated images (photographs derived randomly from a stock corpus) – a modulation that was not observed when garbled images were repeated. For immediately repeated non-target stimuli (novel geometric line drawings), Penney et al. (2003) describe a significant reduction in MEG amplitude in the 250 - 450ms range in parietal/frontal areas and attribute this decline either to reduced attention or to the facilitated processing of the repeated item. Notably, such response suppression for face stimuli was observed by Henson et al. (2002) in the right lateral fusiform area (famous faces) and left inferior occipital area (famous and unfamiliar), but only when the subjects performed an implicit fame-judgment task, not when they performed an explicit repetition detection task. In addition to this fMRI experiment, Henson et al. (2003) combined fMRI and EEG to compare responses to scrambled versus intact face images; face recognition was correlated with an increase in hemodynamic response in fusiform and medial and orbito-frontal cortices and a central, frontal EEG positivity at 550ms, and exposure to a repeated face exhibited an enhanced EEG positivity from 400 - 600ms.

A very closely related question to that of mere repetition concerns the perceptual learning (or perhaps "priming" or "hysteresis," depending on the particular experimental context) that occurs when one's experience with degraded images, whether static or continuous transforming, affects ones later ability to recognize the same items when next encountered (perhaps in an even more corrupted form). An experiment by Dolan et al. (1997) used fMRI to study the change in neural activity when a previously unrecognized degraded images was later correctly recognized, following exposure to the high-quality source images. Coincident with this form of perceptual learning, they reported enhanced inferotemporal and parietal activity for both object and face stimuli, along with an apparent "coupling" of activity between a inferotemporal face-selective area and medial parietal cortex coinciding with successful face perception.

In the present study, we explore the perceptual onset and hysteresis for both face and object stimuli, and do so in a manner that allows for a comparison of neural activity that is both categorical (in terms of perceptual measures) and parametric (in terms of graded levels of image structure disruption). Further, all comparisons are between images or image classes that have been normalized on a number of dimensions previously shown to influence both perceptual and neural measures (e.g., Avidan et al., 2002). More specifically, we ask if, for faces and perhaps even objects, the M100 and M170 components can be seen to exhibit significant modulations as a result of perceptual onset and as a function of repeated viewings – both of the identical image at a later time (item perceived in the second but not first viewing), and of the same object/face but as seen in other images with randomly varied noise. Stated generally, we find significant modulation of the M100 and M170 for both faces and objects at the onset of perception and with repeated viewings. Further, for the M170 we demonstrate an enhancement effect for repeated viewings of face images and a suppression effect for repeated object images.

3.2 Methods

Participants. Seven subjects (aged 18 - 29 years, four female, one left-handed) participated in this study. All subjects were undergraduate or graduate students at the Massachusetts Institute of Technology (MIT) and provided written informed consent, and received payment, according to MIT guidelines. An additional four volunteers were excluded from the study, each according to one of the following criteria: absence of discernible or selective M170, excessive and untimely eye-blink artifacts, neglected orthodontics, and somnolence.

Stimuli. One hundred and twenty RISE sequences were generated using the phase manipulation technique described by Sadr and Sinha (2001, in press), including image preprocessing to normalize the luminance histograms and power spectra across stimulus categories. Of these sequences, sixty were based on images of natural and man-made objects (e.g., a seashell, my father's Zippo®), and sixty were based on images of faces. The latter group was composed of an equal number of familiar (i.e., famous; e.g., Sylvester Stallone, Julia Roberts) and unfamiliar faces, and the objects selected for presentation in the study were chosen not only so as to be common and easily recognized but also to form a fairly heterogenous set, so as to reduce potential effects of cross-object priming. Each of the original source images was a grayscale, photographic quality 255x255 pixel image of the item on a dark background, and for each a series of RISE images was generated at four equally-spaced levels of phase coherence: 65% (worst), 75%, 85%, 95% (best). Degradation levels were based on prior subject testing (e.g., Sadr & Sinha, 2003, in press) in such a manner as to well-span the dynamic range of perceptual onsets and hysteresis (i.e., at least the worst frame should not be recognized at first exposure, at least the best one or two frames should be recognized on first exposure, and there should be at least one level of degradation at which subjects do not perceive the object at first exposure but do so at second exposure). The stimuli were gamma-corrected for presentation via a mirror-reflected projection display and were viewed binocularly by the subject in a dimly lit, magnetically shielded room at a viewing distance of approximately 26 cm ($\sim$12 degrees of visual angle).

Procedure. Each four-frame RISE series was presented as a seven-frame palindromic series – all four images first, in order of worst to best, followed by the first three images again but in reverse order (i.e., 65%, 75%, 85%, 95%, 85%, 75%, 65%; see figure 3-2). Further, in a manner not unlike the random distractor technique described by Sadr and Sinha (2001, in press), each of the 120 RISE sequences was paired and randomly interleaved with another randomly chosen sequence, so ultimately 60 sequences were presented of 14 frames each. That is, referring to the two items in each pair as A and B, each 14-frame sequence might be of the form A-65%, B-65%, ... , B-95%, A-95%, ... , A-65%, B-65%. These random item pairing were counter-balanced for stimulus type (e.g., equal probability of a face sequence being paired with an object sequence as with another face sequence, etc.). The presentation of the 60 paired/interleaved sequences was subdivided into five blocks of twelve, separated by four subject-controlled rest periods. The experiment proper was preceded with two practice runs (four objects) using images/objects that would not be seen again. The subjects' task in the experiment was to press a button in response to any presented image that could not be recognized; while perhaps initially counter-intuitive, this resulted in much less button-pressing on the part of the subjects. Notwithstanding subjects' ability to apprehend the fifty percent probability of an item being either a familiar or unfamiliar face, subjects were naive as to what items (i.e., particular objects or people) would appear in each RISE sequence, and each RISE sequence appeared only once within the experiment.

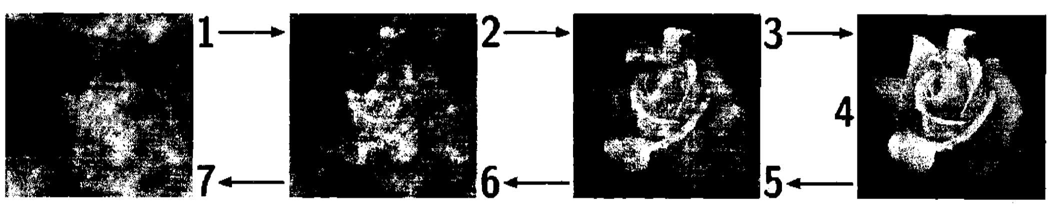


Figure 3-2: Sample RISE sequence depicting four frames at equally spaced levels of degradation. Each four-frame series was converted to a seven-frame sequence for presentation in the experiment and, further, interleaved with another randomly selected seven-frame sequence (see text).

The time-course of the presentation of each frame was as follows (see also figure 3-3). First, a cross was presented at the center of the screen for a duration of 150 - 250ms (uniform random distribution), during which time subjects were to fixate the cross and, thereafter, not move or blink the eyes until the offset of the stimulus image. Second, the stimulus image was presented for 600ms. Third, the image was replaced by a blank square for one second. Subjects could respond with a button press during either image presentation or within the first 800ms of the blank post-/inter-stimulus period, and they were instructed to not be concerned with responding quickly.

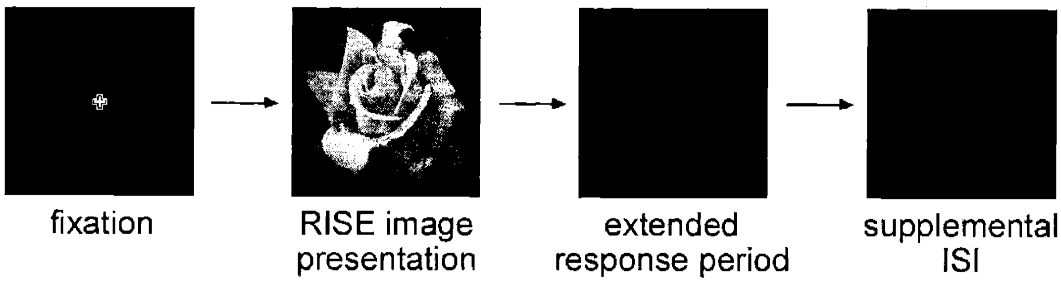


Figure 3-3: Time-course of each presented frame: 150 - 250ms fixation period, 600ms image presentation, 800ms post-presentation response period (note that subjects could also respond during image presentation), 200ms interstimulus interval. Subjects were asked to fixate and not blink or move their eyes during the fixation and image presentation periods.

MEG Recording. Recordings were performed using a supine 160-channel whole-head MEG system with first-order gradiometer sensors. MEG signals were measured continuously (along with triggers sent on eight additional channels for subsequent data processing), digitized at a sampling rate of 500Hz, and pre- and post-amplified by factors of 5 and 50, respectively. At time of recording, signals were band-passed in the range of 1Hz to 200Hz with a notch at 60Hz to remove ambient electrical interference.

MEG Data Processing. Following experiment, noise-reduction was performed on the entire MEG recording based on the measurements taken at three reference sensors (2s moving window), and the data was low-pass filtered at 40Hz (127-point Hamming window). Based on triggers sent from the stimulus computer to the acquisition system 150ms prior to each image onset, the MEG signal corresponding to each image presentation, spanning 100ms pre-image-onset to 500ms post-image-onset, was extracted from the continuous recording. Based on both the experimental protocol and behavioral responses, the trials were sorted by stimulus category (faces versus objects), then further subdivided by the trial's order within its sequence (i.e., trials 1 to 7) and by the subject's perceptual report for that trial. Regarding the latter, the following trial groupings were formed: preonset (last frame when button pressed), onset, onset-complement (the same degradation-level image as the onset, but when seen again in the latter half of the sequence), and preonset-complement (the same degradation-level image as the preonset, but when seen again in the latter half of the sequence). Note that the preonset-complement group included only images for which the subject reported perceiving the item (i.e., same image as preonset, but preonset-complement images were perceived). With each of these trial groupings formed, and following removal of individual trial exhibiting noise artifacts, time-series means and standard deviations were calculated for each group.

Following averaging, a time-series t-test comparison of level 4 (best) and level 1 (worst, including only images which were explicitly reported as not recognized) recordings was performed to isolate the significantly selective ($p < 0.05$) sensors of interest (SOIs) for subsequent analysis of the other categories. The SOI selection was done separately for each subject, each of the two stimulus categories (face and object), and each of the two signal components of interest (M100 and M170). Regarding the latter, the left and right hemisphere SOIs chosen were taken from those occipitotemporal sensor that fit the polarity and latency profile described in previous work (e.g., Linkenkaer-Hansen, et al., 1998; Liu, Harris & Kanwisher, 2002; and others).

Operating on the data recorded at these SOIs, grand means and variances of amplitude peaks and troughs were calculated for each condition, along with grand means and variances of the amplitude differences between each pair of conditions to be compared. The latter measure, although very closely related to the former measure, is less sensitive to systematic individual differences in amplitude (i.e., more robust to the effects of combining data from subjects who produce consistently very large signals and subject who produce much smaller signals). Finally, t-tests were calculated for the peak signal amplitudes (differences) for each of the paired conditions of interest (e.g., preonset versus preonset complement). These t-test results and grand means are reported in the following section.

3.3 Results

Figure 3-4 shows the grand means and standard errors for each of the perceptually-defined trial categories for face stimuli, separately for the M100 and M170 components, while figure 3-5 shows the same for object stimuli. For the M100 component corresponding to face stimuli, the following comparisons yielded significant differences: preonsets trials exhibited significantly smaller amplitudes as compared to preonset-complements ($p < 0.001$, left hemisphere, $p < 0.05$, right hemisphere); preonset trials also exhibited significantly smaller amplitude as compared to onset trials ($p < 0.05$, left hemisphere, $p < 0.001$, right hemisphere); onset trials were significantly smaller in amplitude than onset-complements in the left hemisphere ($p < 0.05$) and significantly larger in the right hemisphere ($p < 0.001$); onset trials also exhibited significantly smaller amplitudes as compared to the level 4 (best image) trials ($p < 0.001$, right and left hemispheres). Onset amplitudes for familiar faces did not differ significantly with those for unfamiliar faces (combined in figure 3-4). M100 latency was significantly earlier for preonsets than for preonset-complements in the left hemisphere ($p < 0.05$) and significantly earlier for onsets than for level 4 trials in the right ($p < 0.05$).

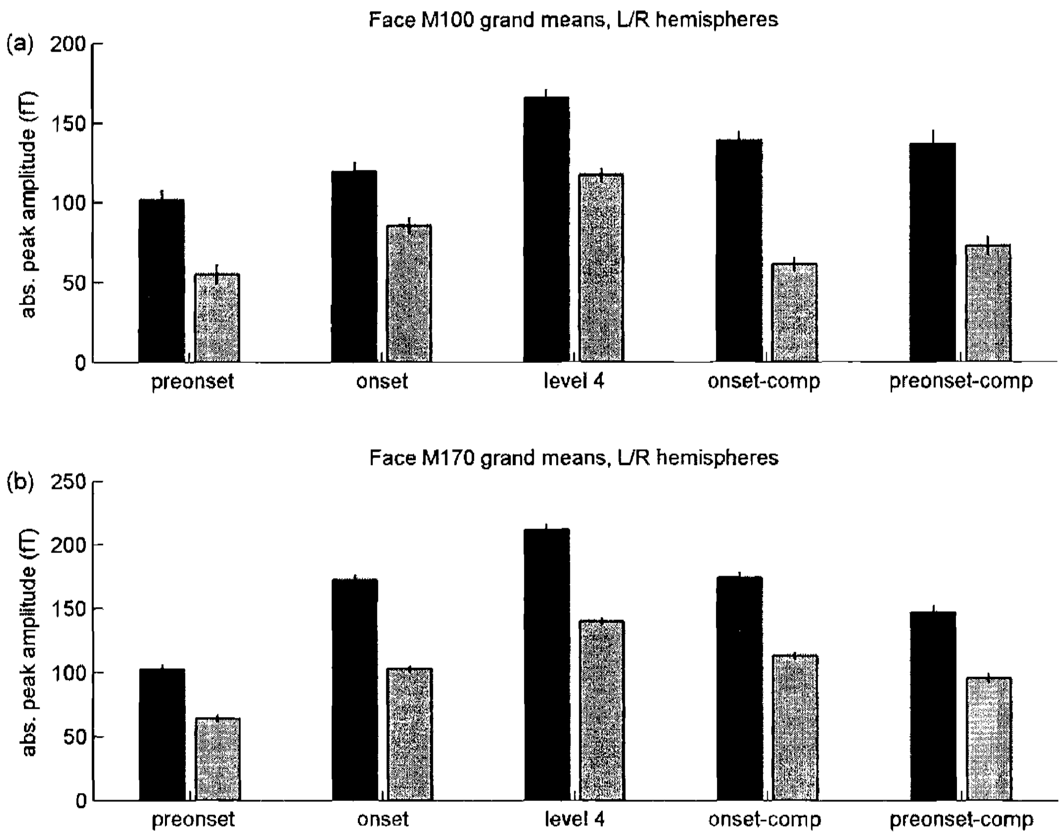


Figure 3-4: Grand means and standard errors of peak amplitudes for each perceptually-based trial category, face stimuli. The results consistently show preonsets < preonset-complements, preonsets < onsets, and onsets < level 4 (best images). M170 measures also indicate a neural enhancement effect for face stimuli (e.g., compare onsets and onset-complements, right hemisphere). (a) M100 troughs (left hemisphere, dark gray, absolute value shown) and peaks (right hemisphere, light gray). (b) M170 peaks (left hemisphere, dark gray) and troughs (right hemisphere, light gray, absolute value shown). See text for statistical tests between conditions.

For the M170 component corresponding to face stimuli, the following comparisons yielded significant differences: preonsets trials exhibited significantly smaller amplitudes as compared to preonset-complements ($p < 0.001$, left and right hemispheres); preonset trials also exhibited significantly smaller amplitude as compared to onset trials ($p < 0.001$, left and right hemispheres); onset trials were significantly smaller in amplitude than onset-complements in the right hemisphere ($p < 0.05$), with $p = 0.0529$ for both hemispheres combined; onset trials also exhibited significantly smaller amplitudes as compared to the level 4 (best image) trials ($p < 0.001$, right and left hemispheres); onset amplitudes for familiar faces were significantly larger than for unfamiliar faces in the left hemisphere (p

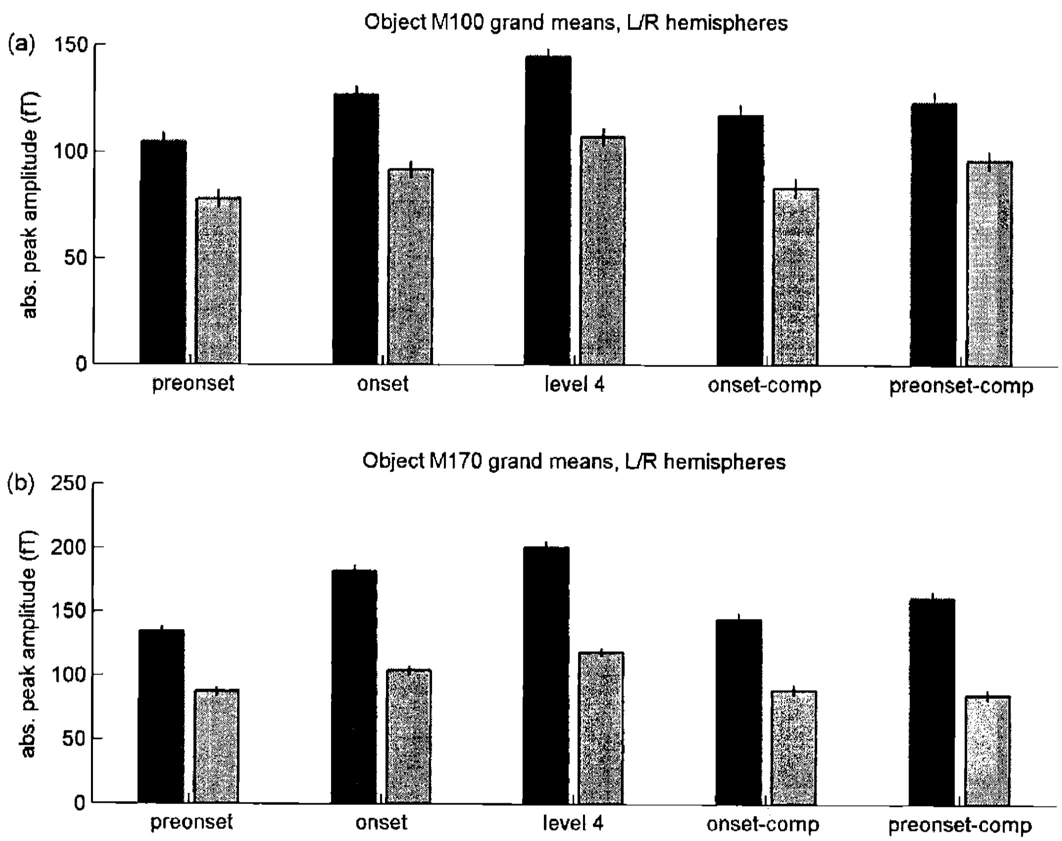


Figure 3-5: Grand means and standard errors of peak amplitudes for each perceptually-based trial category, object stimuli. The results consistently show preonsets < preonset-complements, preonsets < onsets, and onsets < level 4 (best images). M170 measures also indicate a neural suppression effect for object stimuli (e.g., compare onsets and onset-complements). (a) M100 troughs (left hemisphere, dark gray, absolute value shown) and peaks (right hemisphere, light gray). (b) M170 peaks (left hemisphere, dark gray) and troughs (right hemisphere, light gray, absolute value shown). See text for statistical tests between conditions.

< 0.05) and significantly smaller in the right hemisphere ($p < 0.05$); M170 latency was significantly later for onsets than for level 4 trials in the left hemisphere ($p < 0.05$) and significantly earlier for onsets than for onset-complements in the right ($p < 0.01$).

For the M100 component corresponding to object stimuli, the following comparisons yielded significant differences: preonsets trials exhibited significantly smaller amplitudes as compared to preonset-complements ($p < 0.01$, left and right hemispheres); preonset trials also exhibited significantly smaller amplitude as compared to onset trials ($p < 0.001$, left hemisphere, $p < 0.01$, right hemisphere); onset trials exhibited significantly smaller amplitudes as compared to the level 4 (best image) trials ($p < 0.01$, right and left hemispheres).

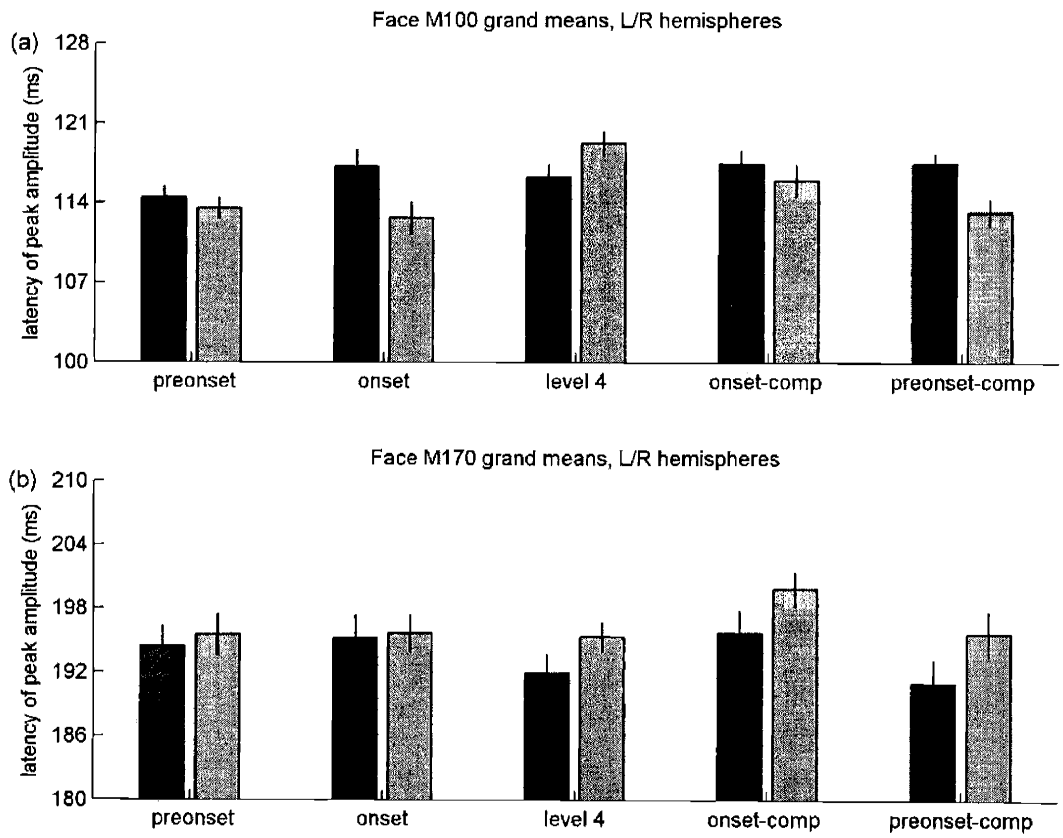


Figure 3-6: Grand means and standard errors of the latencies of peak amplitudes for each perceptually-based trial category, face stimuli. Please see text for results of statistical comparisons between conditions.

For the M170 component corresponding to object stimuli, the following comparisons yielded significant differences: preonsets trials exhibited significantly smaller amplitudes as compared to preonset-complements in the left hemisphere ($p < 0.001$) and in both hemispheres combined ($p < 0.01$) but not in the right hemisphere alone; preonset trials exhibited significantly smaller amplitude as compared to onset trials ($p < 0.001$, left and right hemispheres); onset trials were significantly smaller in amplitude than onset-complements in the right hemisphere ($p < 0.05$) and for both hemispheres combined ($p < 0.05$) but not significantly smaller in the left hemisphere alone; onset trials also exhibited significantly smaller amplitudes as compared to the level 4 (best image) trials ($p < 0.001$, left hemisphere, $p < 0.01$, right hemisphere).

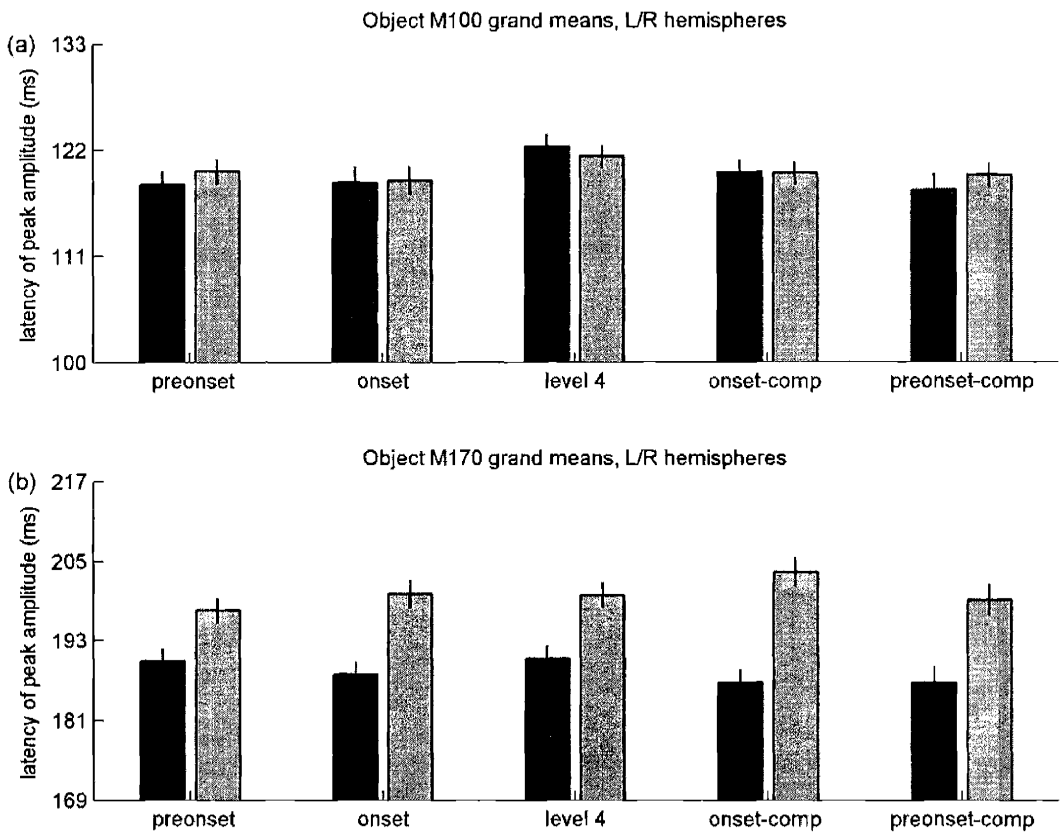


Figure 3-7: Grand means and standard errors of the latencies of peak amplitudes for each perceptually-based trial category, object stimuli. While suggestive of an earlier M100 latency for the level 4 (best) images, no comparisons here yielded significant differences (except that the M170 tends to peak earlier in the left hemisphere as compared to the right).

3.4 Discussion

In this experiment, we characterize explicit object and face perception using RISE image sequences in order to better index the onset and offset of perception without concern for the confounding covariance of a number of low-level image properties. One key benefit is the ability to compare neural responses to a given degraded image under different perceptual conditions, but beyond comparing identical stimuli under different perceptual conditions, the comparison of responses to the (physically different) pre- and post-onset images is bolstered by the tight control of important image properties, allowing an interpretation of the findings more closely tied to the evolution of meaningful structure of the image and to shifts in subjective percepts. In this manner, one may investigate the rise and fall of neural activity frame by frame, as their associated perceptual phenomena manifest themselves. What follows is a more detailed discussion of some of the key findings listed above, their theoretical and methodological implications, and the manner in which they may speak to some closely related work in the field.

Perhaps the most consistent and fundamental finding here is that the preonset trials showed M100 and M170 amplitudes that were smaller than those of the onset trials and, even more strikingly, the preonset-complement trials. The comparison of the preonset to onset shift is nicely illustrative of the reflection of the perceptual change in the neural signal, and the comparison to the preonset complement confirms that this increase in amplitude was not likely due simply to the stimulus difference between the preonset and onset degradation levels. Again, as with such work as by Dolan et al (1997), the preonset/preonset-complement comparison is between identical images, with the only difference being that of the subjects' percepts (item not perceived when seen at preonset but perceived when same image was presented again in the latter half of the sequence). Notably, such a comparison of the MEG data is made possible by the marked hysteresis (or "perceptual learning") effect previously described for these kinds of stimuli and presentation approaches (e.g., Sadr & Sinha, in press). This comparative approach resembles, of course, recent work by Liu, Harris, and Kanwisher (2002) using similar RISE stimuli of faces and houses, presented in randomized fashion at previously established threshold levels of degradation. A notable difference between the two studies, however, is the mode of presentation (sequential versus random), combined with the lack of repetition of RISE sequences, which allows the further analysis of frame-by-frame changes in neural activity (especially as discussed further below), not only a difference between perceived and not perceived but otherwise identical stimuli. Another important point is that, given the lack of modulation of the M100 response by the successful perception of houses, one may have predicted that the M100 component is similarly not affected by the perception of objects; however, we have shown here the contrary. (One way this house/object discrepancy could be reconciled is with reference to work by Epstein and Kanwisher (1999) suggesting that house perception, and, in general, place perception, may well be expected to show these sorts of very early perception-induced effects but at a different cortical location and, thus, under sensors of interest different from those selected for analysis in the above study.) Finally, it is worth noting that, as compared to a number of studies that have explored the effects of stimulus repetition (discussed above), the current technique affords one the ability to present an image twice and still have control over whether the item depicted therein is, in fact, perceived twice (e.g., onset versus onset-complement comparison) or only once (e.g., preonset to preonset-complement comparison), thereby discriminating between mere repetition effects that may be very low-level in nature and not tied to higher-level percepts.

Closely related to this issue is the rather interesting finding of M170 repetition response enhancement for face stimuli, along with a contrary suppression of response for object stimuli, as seen in the comparison of onset versus onset-complement trials. First, this finding provides even more support for the notion that the perceptual-hysteresis-based neural signal differences discussed here are not the result of a low-level, mere repetition effect, given that the nature of the stimulus was crucial in determining whether the neural response at later viewings would be amplified or reduced. Second, this result speaks, in a rather interesting way, to the face-selectivity of the M170 response component. Specifically, although, as we have shown here, the M170 can indeed be seen to demonstrate significant modulation coincident with object perception (even though the absolute amplitudes for objects may be smaller than those for faces), it does appear that substantial face-selectivity of the M170 nevertheless rears a less welcoming head at object stimuli at subsequent repeated processing. This effect is in a way not unlike the finding reported by Penney et al. (2003), wherein non-target geometric line-drawing elicited reduced response on immediately repeated presentation. In contrast, however, the response suppression seen in the present study seems much less likely to be tied to explicit task demands (i.e., having been labeled as distractors in the first place) or to the relatively non-evocative nature of the stimuli (line drawings of manufactured geometric shapes) of the experiment by Penney et al. Rather, the response suppression seen for objects in this study seems to suggest an inherent bias of the underlying processing mechanism toward faces and against objects, even if, on their first being perceived, objects do elicit a modest response. Moreover, even as the enhancement/suppression effect here of course speaks to a growing neuroimaging literature interested in the effects of repetition and priming on hemodynamic and neural signals (e.g., Henson et al., 2003; Halgren et al., 2000; Debruille, Guillem & Renault, 1998; and others), this result also speaks to an important technical feature of the current study – specifically, that the analysis of responses to objects (beginning with the selection of the most informative sensors) was conducted independently of face stimuli and for its own sake, not simply as a subtractive baseline by which to characterize the face-selective responses. Finally, it is interesting to see another methodological prescription in these findings: that the random but repeated presentation of a collection of images for the purposes to characterizing the response to, for example, different stimulus classes. Unless one intends to study response enhancement/suppression due to cross-trial priming and/or repetition, one would do well to not repeatedly present the same or very similar stimuli, nor images that could be familiar *a priori*. In the specific case of M170 responses to faces versus objects, such stimulus presentation should be expected to expand the gulf between the response magnitudes to the object classes.

Recent work by Allison, Puce, and McCarthy (2002) provides some interesting hints as to the neural mechanisms that may underlie such a category-selective repetition enhancement/suppression effect. In this study, at approximately half of the occipitotemporal cortical sites characterized by electrocorticography as producing face- or word-selective N200 responses, dispreferred stimuli elicited not merely a smaller N200 response but rather a positivity ("P200"). These authors go on to interpret these results via a relatively simple model of local lateral interactions in which these observed P200 components are purported to coincide with hyperpolarizing inhibition of the apical dendrites of pyramidal cells, as applied by nearby cells with contradictory stimulus-class selectivity. Conversely, the basic N200 component observed in these fusiform and inferotemporal sites is interpreted as the product of excitatory depolarizing potentials applied by nearby cells of congruous selectivity. One could well imagine that such a mechanism, amplified in some manner by repeated stimulus presentations, could also serve to produce a category-selective M170 repetition enhancement/suppression effect such as has been observed in the current study.

Our purpose in the current paper has been to address a number of technical and conceptual questions relating to behavioral and electrophysiological study of human visual object and face perception. It is hoped not only that the current findings may shed a modest light on certain basic questions surrounding the mechanisms that underlie the onset of object and face perception, their modulation by perceptual experience, etc., but also that the nature of the experimental approach itself may in some way present a worthwhile contribution in its own right.

Some say the world will end in fire,
some say in ice.
From what I've tasted of desire
I hold with those who favor fire.
But if it had to perish twice,
I think I know enough of hate
to say that for destruction ice
is also great
and would suffice.

- Robert Frost -

Chapter 4

The Fidelity of Local Ordinal Encoding

4.1 Introduction

Biological and artificial recognition systems face the challenge of grouping together differing proximal stimuli arising from the same underlying object. How well a system succeeds in overcoming this challenge is critically dependent on the nature of the internal representations against which the observed inputs are matched. The representation schemes should be capable of efficiently encoding object concepts while tolerating variations in appearance.

In this paper, we introduce and characterize a biologically plausible representation scheme for encoding signal structure. The scheme employs a simple vocabulary of local ordinal relations, of the kind that early sensory neurons are capable of extracting. Our results so far suggest that this scheme possesses several desirable characteristics, including tolerance to object appearance variations, computational simplicity, and low memory requirements. We develop and demonstrate our ideas in the visual domain, but they are intended to be applicable to other sensory modalities as well.

The starting point for our proposal lies in studies of the response properties of neurons in the early sensory cortical areas. These response properties constrain the kinds of measurements that can plausibly be included in our representation scheme. In the visual domain, many striate cortical neurons have rapidly saturating contrast response functions [4, 22]. Their tendency to reach ceiling level responses at low contrast values render these neurons sensitive primarily to local ordinal, rather than metric, relations. We propose to use an idealization of such units as the basic vocabulary of our representation scheme (figure 4-1). In this scheme, objects are encoded as sets of local ordinal relations across image regions. As discussed below, this very simple idea seems well suited to handling the photometric appearance variations that real-world objects exhibit.

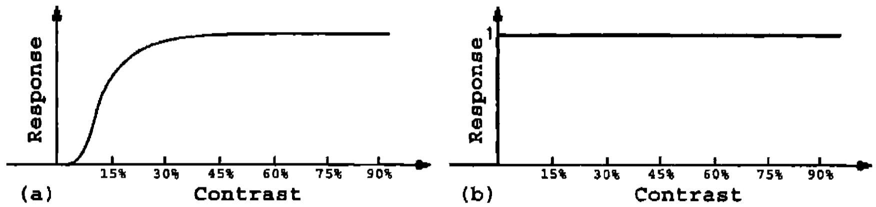


Figure 4-1: (a) A schematic contrast response curve for a primary visual cortex neuron. The response of the neuron saturates at low contrast values. (b) An idealization of (a). This unit can be thought of as an ordinal comparator, providing information only about contrast polarity but not its magnitude.

As shown in figure 4-2, variations in illumination significantly alter the individual brightness of different parts of the face, such as the eyes, cheeks, and forehead. Therefore, absolute image brightness distributions are unlikely to be adequate for classifying all of these images as depicting the same underlying object. Even the contrast magnitudes across different parts of the face change greatly under different lighting conditions. While the absolute luminance and contrast magnitude information is highly variable across these images, Thoresz and Sinha [90] have shown that one can identify some stable ordinal measurements.

Figure 4-2: The challenge for a representation scheme: to construct stable descriptions of objects despite radical changes in appearance.

Figure 4-3 shows several pairs of average brightness values over localized patches for each of the three images included in figure 4-2. Certain regularities are apparent. For instance, the average brightness of the left eye is always less than that of the forehead, irrespective of the lighting conditions. The relative magnitudes of the two brightness values may change, but the sign of the inequality does not. In other words, the ordinal relationship between the average brightnesses of the {left-eye, forehead} pair is invariant under lighting changes. Figure 4-3 shows several other such pair-wise invariances. It seems, therefore that local ordinal relations may encode the stable facial attributes across different illumination conditions. An additional advantage to using ordinal relations is their natural robustness to sensor noise. Thus, it would seem that local ordinal representations may be well suited for devising compact representations, robust against large photometric variations, for at least some classes of objects. Notably, for similar reasons, ordinal measures have also been shown to be a powerful tool for simple, efficient, and robust stereo image matching [10].

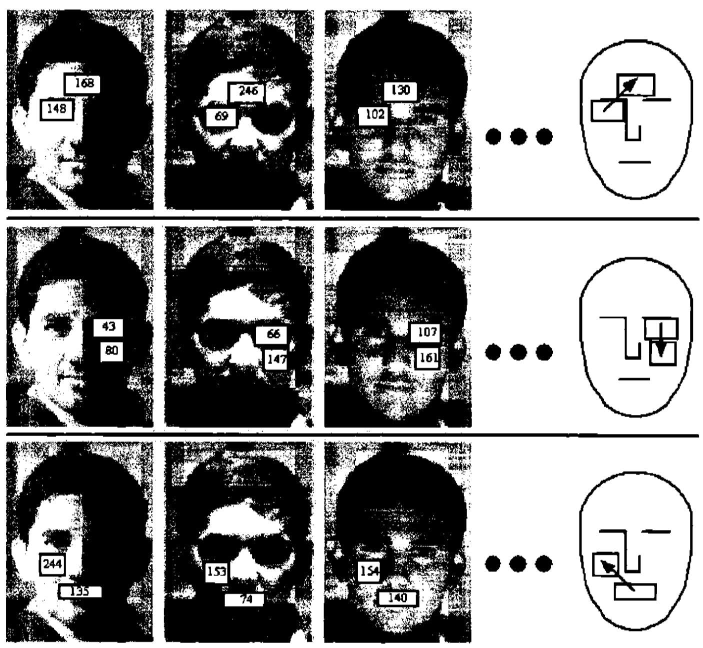


Figure 4-3: The absolute brightnesses and their relative magnitudes change under different lighting conditions but several pair-wise ordinal relationships stay invariant.

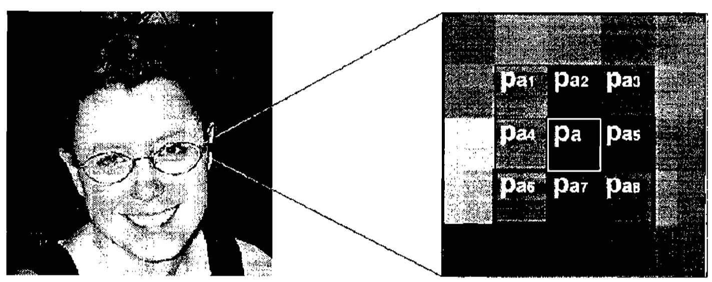
In what follows, we address an important open question regarding the expressiveness of the ordinal representation scheme. Given that this scheme ignores absolute luminance and contrast magnitude information, an obvious question that arises is whether such a crude representation strategy can encode object/image structure with any fidelity.

4.2 Information Content of Local Ordinal Encoding

Figure 4-4 shows how we define ordinal relations between an image region p_a and its immediate neighbors $p_b = \{p_{a1}, \ldots, p_{a8}\}$. In the conventional rectilinear grid, when all image regions p_a are considered, four of the eight relations are redundant; we encode the remaining four as $\{1, 0, -1\}$ based on the difference in luminance between two neighbors being positive, zero, or negative, respectively. To demonstrate the richness of information encoded by this scheme, we compare the original image to one produced by a function that reconstructs the image using local ordinal relationships as constraints. Our reconstruction function has the form

$$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}), \tag{4.1}$$

where $\mathbf{x} = \{i, j\}$ is the position of a pixel, $f(\mathbf{x})$ is its intensity, ϕ is a map from the input space into a high (possibly infinite) dimensional space, $\mathbf{w}$ is a hyperplane in this high-dimensional space, and $\mathbf{u} \cdot \mathbf{v}$ denotes an inner product.



$$\begin{array}{rcl} \mathbf{I}(p_a) & < & \mathbf{I}(p_{a_1}) \\ & = & \mathbf{I}(p_{a_2}) \\ & < & \mathbf{I}(p_{a_3}) \\ & < & \mathbf{I}(p_{a_4}) \\ & > & \mathbf{I}(p_{a_5}) \\ & < & \mathbf{I}(p_{a_6}) \\ & < & \mathbf{I}(p_{a_7}) \\ & < & \mathbf{I}(p_{a_8}) \end{array}$$

Figure 4-4: Ordinal relationships between an image region p_a and its neighbors.

Infinitely many reconstruction functions could satisfy the given ordinal constraints. To make the problem well-posed we regularize [91] the reconstruction function subject to the ordinal constraints, as done in ordinal regression for ranking document retrieval results [39]. Our regularization term is a norm in a Reproducing Kernel Hilbert Space (RKHS) [5, 92]. Minimizing the norm in a RKHS subject to the ordinal constraints corresponds to the following convex constrained quadratic optimization problem:

$$\min_{\xi, \mathbf{w}} \frac{1}{2} ||\mathbf{w}||^2 + C \sum_p \xi_p \tag{4.2}$$

subject to

$$\theta(\delta_p) \mathbf{w} \cdot (\phi(\mathbf{x}_{p_a}) - \phi(\mathbf{x}_{p_b})) \geq |\delta_p| - \xi_p, \quad \forall p \text{ and } \xi \geq 0, \tag{4.3}$$

where the function $\theta(y) = +1$ for $y \geq 0$ and -1 otherwise, p is the index over all pairwise ordinal relations between all pixels p_a and their local neighbors p_b (as depicted in figure 4-4), ξ_p are slack variables which are penalized by C (the trade-off between smoothness and ordinal constraints), and δ_p take integer values $\{-1, 0, 1\}$ denoting the ordinal relation (less than, equal to, or greater than, respectively) between p_a and p_b; for the case $\delta_p = 0$ the inequality in (4.3) will be a strict equality.

Taking the dual of (4.2) subject to constraints (4.3) results in the following convex quadratic optimization problem which has only box constraints:

$$\max_{\alpha} \sum_{p} |\delta_p| \, \alpha_p - \frac{1}{2} \sum_{p} \sum_{q} \alpha_p \alpha_q \tilde{\mathbf{K}}_{pq} \tag{4.4}$$

subject to

$$\begin{aligned} 0 \leq \alpha_p \leq C \quad & \text{if} \quad \delta_p > 0, \\ -C \leq \alpha_p \leq C \quad & \text{if} \quad \delta_p = 0, \\ -C \leq \alpha_p \leq 0 \quad & \text{if} \quad \delta_p < 0, \end{aligned} \tag{4.5}$$

where α_p are the dual Lagrange multipliers, and the elements of the matrix $\tilde{\mathbf{K}}$ have the form

$$\begin{aligned} \tilde{K}_{pq} &= (\phi(\mathbf{x}_{p_a}) - \phi(\mathbf{x}_{p_b})) \cdot (\phi(\mathbf{x}_{q_a}) - \phi(\mathbf{x}_{q_b})) \\ &= K(\mathbf{x}_{p_a}, \mathbf{x}_{q_a}) - K(\mathbf{x}_{p_b}, \mathbf{x}_{q_a}) - K(\mathbf{x}_{p_a}, \mathbf{x}_{q_b}) + K(\mathbf{x}_{p_b}, \mathbf{x}_{q_b}), \end{aligned}$$

where $K(\mathbf{y}, \mathbf{x}) = \phi(\mathbf{y}) \cdot \phi(\mathbf{x})$ using the standard kernel trick [66]. In this paper we use only Gaussian kernels $K(\mathbf{y}, \mathbf{x}) = \exp(-||\mathbf{x} - \mathbf{y}||^2/2\sigma^2)$. The reconstruction function, $f(\mathbf{x})$, obtained from optimizing (4.4) subject to box constraints (4.5) has the following form

$$f(\mathbf{x}) = \sum_{p} \alpha_p \left(K(\mathbf{x}, \mathbf{x}_{p_a}) - K(\mathbf{x}, \mathbf{x}_{p_b}) \right). \tag{4.6}$$

Note that in general many of the α_p values may be zero – these terms do not contribute to the reconstruction, and the corresponding constraints in (4.3) were not required. The remaining α_p with absolute value less than C satisfy the inequality constraints in (4.3), whereas those with absolute value at C violate them.

Figure 4-5 depicts two typical reconstructions performed by this algorithm. The difference images and error histograms suggests that the reconstructions closely match the source images.

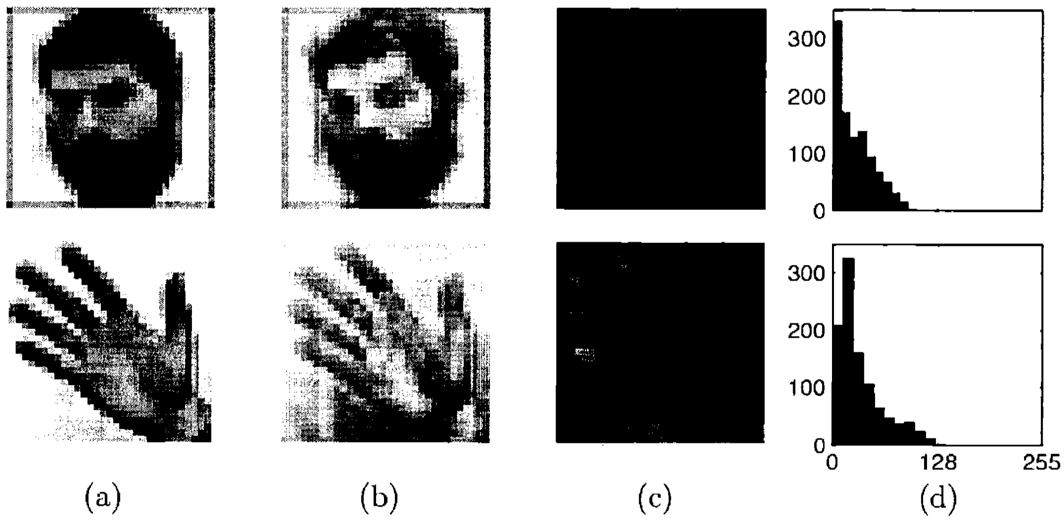


Figure 4-5: Reconstruction results from the regularization approach. (a) Original images. (b) Reconstructed images. (c) Absolute difference between original and reconstruction. (d) Histogram of absolute difference.

4.3 Discussion

Our reconstruction results suggest that the local ordinal representation can faithfully encode image structure. Thus, even though individual ordinal relations are insensitive to absolute luminance or contrast magnitude, a set of such relations implicitly encodes metric information. In the context of the human visual system, this result suggests that the rapidly saturating contrast response functions of the early visual neurons do not significantly hinder their ability to convey accurate image information to subsequent cortical stages.

An important question that arises here is what are the strengths and limitations of local ordinal encoding. The first key limitation is that for any choice of neighborhood size over which ordinal relations are extracted, there are classes of images for which the local ordinal representation will be unable to encode the metric structure. For a neighborhood of size n, an image with regions of different luminance embedded in a uniform background and mutually separated by a distance greater than n would constitute such an image. In general, sparse images present a problem for this representation scheme, as might foveal or cortical "magnification," for example. This issue could be addressed by using ordinal relations across multiple scales, perhaps in an adaptive way that varies with the smoothness or sparseness of the stimulus.

Second, the regularization approach above seems biologically implausible. Our intent in using this approach for reconstructions was to show via well-understood theoretical tools the richness of information that local ordinal representations provide. In order to address the neural plausibility requirement, we have developed a simple relaxation-based approach with purely local update rules of the kind that can easily be implemented by cortical circuitry. Each unit communicates only with its immediate neighbors and modifies its value incrementally up or down (starting from an arbitrary state) depending on the number of ordinal relations in the positive or negative direction. This computation is performed iteratively until the network settles to an equilibrium state. The update rule can be formally stated as

$$\mathbf{R}_{p_a,t+1} = \mathbf{R}_{p_a,t} + \Delta \sum_{p_b} (\theta(\mathbf{R}_{p_a,t} - \mathbf{R}_{p_b,t}) - \theta(\mathbf{I}_{p_a} - \mathbf{I}_{p_b})), \tag{4.7}$$

where $\mathbf{R}_{p_a,t}$ is the intensity of the reconstructed pixel p_a at step t, $\mathbf{I}_{p_a}$ is the intensity of the corresponding pixel in the original image, Δ is a positive update rate, and θ and p_b are as described above. Figure 4-6 shows four examples of image reconstructions performed using a relaxation-based approach.

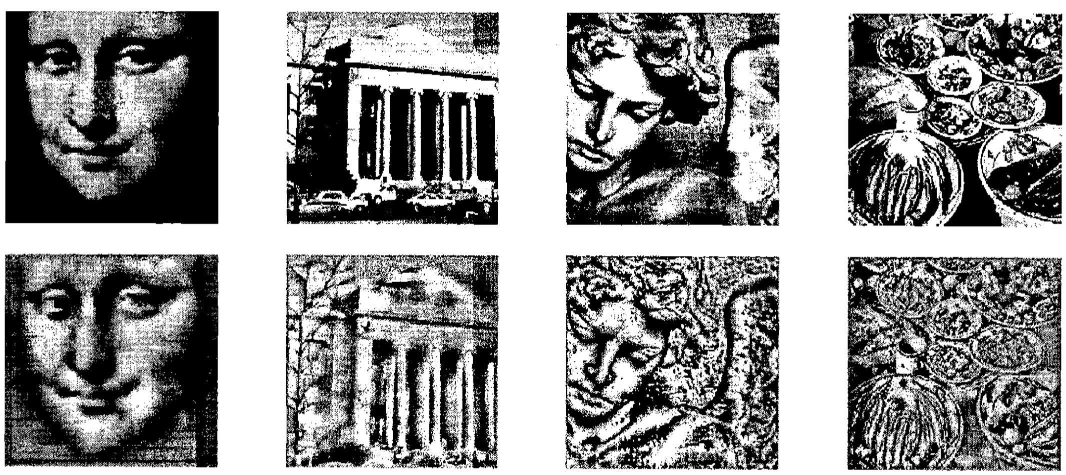


Figure 4-6: Reconstruction results from the relaxation approach.

A third potential limitation is that the scheme does not appear to constitute a compact code. If each pixel must be encoded in terms of its relations with all of its eight neighbors, where each relation takes one of three values, $\{-1, 0, 1\}$, then what has been gained over the original image where each pixel is encoded by 8 bits? There are three ways to address this question.

1. Eight relations per pixel is highly redundant – four are sufficient. In fact, as shown in figure 4-7, the scheme can also tolerate several missing relations.

100% 80% 60% 40% 20%

Figure 4-7: Five reconstructions, shown here to demonstrate the robustness of local ordinal encoding to missing inputs. From left to right: reconstructions based on 100%, 80%, 60%, 40%, and 20% of the full set of immediate neighbor relations.

2. An advantage to using ordinal relations is that they can be extracted and transmitted much more reliably than metric ones. These relations share the same spirit as loss functions used in robust statistics [40] and trimmed or Winsorized estimators.

3. The intent of the visual system is often not to encode/reconstruct images with perfect fidelity, but rather to encode the most stable characteristics that can aid in classification. In this context, a few ordinal relations may suffice for encoding objects reliably. Figure 4-8 shows the results of using less than 20 relations for detecting faces. Clearly, such a small set would not be sufficient for reconstructions, but it works well for classification. Its generalization arises because it defines an equivalence class of patterns.

In summary, the ordinal representation scheme provides a neurally plausible strategy for encoding signal structure. While in this paper we focus on demonstrating the fidelity of this scheme, we believe that its true strength lies in defining equivalence classes of patterns enabling generalizations over appearance variations in objects. Several interesting directions remain to be explored. These include the study of ordinal representations across multiple scales, learning schemes for identifying subsets of ordinal relations consistent across different instances of an object, and the relationship of this work to multi-dimensional scaling [95] and to the use of truncated, quantized wavelet coefficients as "signatures" for fast, multiresolution image querying [42].

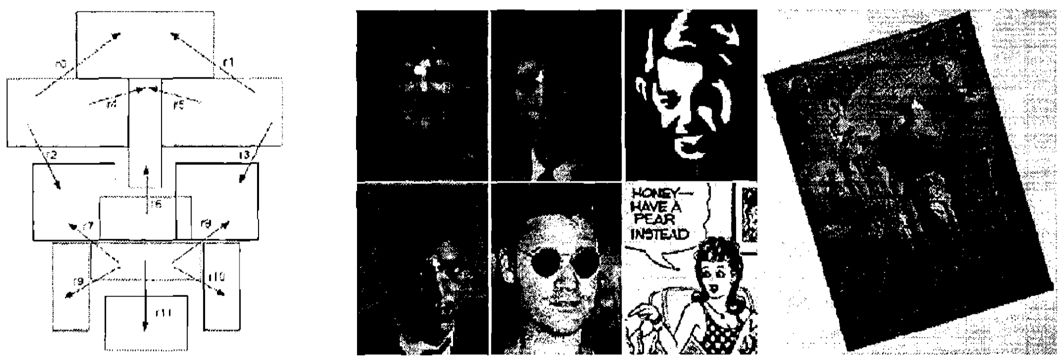


Figure 4-8: A small collection of ordinal relations (a), though insufficient for high fidelity reconstruction, is very effective for pattern classification despite significant appearance variations. (b) Results of using a local ordinal relationship based template to detect face patterns. The program places white dots at the centers of patches classified as faces. (From Thoresz and Sinha, in preparation.)

Others taunt me with having knelt at well-curbs
always wrong to the light, so never seeing
deeper down in the well than where the water
gives me back in a shining surface picture
my myself in the summer heaven, godlike,
looking out of a wreath of fern and cloud puffs.
Once, when trying with chin agains a well-curb,
I discerned, as I thought, beyond the picture,
through the picture, a something white, uncertain,
something more of the depths – and then I lost it.
Water came to rebuke the too clear water.
One drop fell from a fern, and lo, a ripple
shook whatever it was lay there at bottom,
blurred it, blotted it out. What was that whiteness?
Truth? A pebble of quartz? For once, then, something.

- Robert Frost -

Chapter 5

Conclusion

5.1 Contributions

In closing we briefly summarize here some of the main theoretical, methodological, and empirical contributions of the work included in this thesis followed, in the following section, by a short survey of interesting avenues of further research.

In many ways the primary contribution of this work is the development of a flexible image processing and stimulus presentation paradigm (RISE) for the experimental study of a number of key issues in high-level vision. Included in this endeavor was the experimental use of this new paradigm to characterize the onset of conscious visual perception in impoverished settings, perceptual hysteresis during progressive image degradation, and the effect of semantic-perceptual priming on the threshold of object recognition.

The second major product of this work is the extension of RISE into the realm of brain imaging/electrophysiology, and the application this new combined methodology to the study of specific issues surrounding the neural substrates of object and face recognition. In particular, this includes the characterization of the neural correlates of conscious perceptual onset along RISE image trajectories and a comparison of neural activity associated with the perception of differing object classes (faces and objects). An intriguing and frankly unexpected finding associated with the latter is the discovery of a contrary neural response enhancement/suppression effect corresponding to repeated exposures to items from these two classes of objects.

Finally, we have here developed and characterized a simple but robust representation scheme for coding 2D visual signals as simple sets of local ordinal relations and have verified the representational fidelity of this admittedly lossy coding scheme by evaluating its resultant image reconstructions. Included in this endeavor was the development of appropriate reconstruction techniques, notably one that could conceivably be quite readily implemented in neural hardware.

5.2 Future Directions

Then intent of the work documented in this thesis has, in each case, not been to declare "case closed" on a given neuroscientific or methodological concern but rather to open doors to new approaches and inquiries. In this spirit, we briefly consider here a small set of future directions for research related to specific issues raised by the work completed thus far.

Regarding the RISE paradigm in general, one may readily conceive of a great deal of future experiments, from those addressing in more detail the issues of priming, perceptual learning, and hysteresis, to those which bring the technique into research domains as yet untapped by this author. Of these latter, some of the most interesting may be those in the clinical setting (e.g., the development of diagnostic tools for agnosia and prosopagnosia) and in the realm of visual/cognitive development (e.g., the progression of face/object perceptual ability with age).

The promise of RISE in the realm neuroimaging and electrophysiology has already begun to be realized, not only by the work described herein but by other researchers that have taken on the technique for their own purposes (see chapter 3 for details). A great deal remains to be done in this sphere, however, not the least of which is the translation of this approach into neuroimaging experiments which may better localize the sources of neural activity associated with the magnetoencephalographic signal modulation described here.

Finally, in terms of further work on our modeling endeavors concerned with the development and assessment of simple ordinal codes for biologically plausible image representation, it is satisfying to report on a closely related project currently underway in the lab, with the interesting elaboration of the technique to include non-local ordinal measures in encoding the structure of an image at different scales.

... I should not be withheld but that some day
into their vastness I should steal away ...

... They would not find me changed from him they knew –
only more sure of all I thought was true.

\- Robert Frost -

BIBTEX has a way of crushing a man.

\- Javid Sadr -

Bibliography

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

[2] T. Allison, A. Puce, and G. McCarthy. Category-sensitive excitatory and inhibitory processes in human extrastriate cortex. *J. Neurophysiology*, 88:2864–2868, 2002.

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

[4] A. Anzai, M. A. Bearse, R. D. Freeman, and D. Cai. Contrast coding by cells in the cat's striate cortex: Monocular vs. binocular detection. *Visual Neuroscience*, 12:77–93, 1995.

[5] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 686:337–404, 1950.

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

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

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

[9] P. J. Benson and D. I. Perrett. Extracting prototypical facial images from exemplars. *Perception*, 22:257–262, 1993.

[10] D. Bhat and S. Nayar. Ordinal measures for image correspondence. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 351–357, 1996.

[11] R. N. Bracewell. *The Fourier Transform and Its Applications.* McGraw Hill, Boston, 2000.

[12] R. M. Brooks and A. G. Goldstein. Recognition by children of inverted photographs of faces. *Child Development*, 34:1033–1040, 1963.

[13] J. S. Bruner and M. C. Potter. Interference in visual recognition. *Science*, 144:424–425, 1964.

[14] R. L. Buckner, P. A. Bandettini, K. M. O'Craven, R. L. Savoy, S. E. Petersen, M. E. Raichle, and B. R. Rosen. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA*, 93:14878–14883, 1996.

[15] R. L. Buckner, J. Goodman, M. Burock, M. Rotte, W. Koutstaal, D. Schacter, B. Rosen, and A. M. Dale. Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron*, 20(2):285–296, 1998.

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

[17] S. Carey and R. Diamond. From piecemeal to configurational representation of faces. *Science*, 195:312–314, 1977.

[18] S. Carey and R. Diamond. Are faces perceived as configurations more by adults than by children? *Visual Cognition*, 213:253–274, 1994.

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

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

[21] A. Damasio, H. Damasio, and G. van Hoessen. Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology*, 32:331–341, 1982.

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

[23] R. Desimone, T. D. Albright, C. G. Gross, and C. Bruce. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neuroscience*, 4:2051–2062, 1984.

[24] R. L. DeValois and K. K. DeValois. *Spatial Vision*. Oxford Univ. Press, New York, 1988.

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

[26] R. J. Dolan, G. R. Fink, E. Rolls, M. Booth, A. Holmes, R. S. J. Frackowiak, and K. J. Friston. How the brain learns to see objects and faces in an impoverished context. *Nature*, 389:596–599, 1997.

[27] S. Edelman. Representation of similarity in 3D object discrimination. *Neural Computation*, 7:407–422, 1995.

[28] D. C. Van Essen and E. A. DeYoe. Concurrent processing in primate visual cortex. In M. S. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 383–400, Cambridge, MA, 1995.

[29] M. J. Farah. *Visual Agnosia: Disorders of Object Recognition and What They Tell Us about Normal Vision*. MIT Press, Cambridge, MA, 1990.

[30] R. M. French, M. Mermillod, P. Quinn, A. Chauvin, and D. Mareschal. 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*, 2002.

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

[32] E. S. Gollin. Developmental studies of visual recognition of incomplete objects. *Perceptual and Motor Skills*, 11:289–298, 1960.

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

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

[35] E. Halgren, T. Raij, K. Marinkovic, V. Jousmäki, and R. Hari. Cognitive response profile of the human fusiform face area as determined by MEG. *Cerebral Cortex*, 10:69–81, 2000.

[36] L. D. Harmon and B. Julesz. Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, 180(91):1194–1197, 1973.

[37] R. N. Henson, Y. Goshen-Gotstein, T. Ganel, L. J. Otten, A. Quayle, and M. D. Rugg. Electrophysiological and haemodynamic correlates of perception, recognition, and priming. *Cerebral Cortex*, 13:793–805, 2003.

[38] R. N. Henson, T. Shallice, M. L. Gorno-Tempini, and R. J. Dolan. Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex*, 12:178–186, 2002.

[39] R. Herbrich, T. Graepel, and K. Obermeyer. Support vector learning for ordinal regression. In *Proc. of the Ninth Intl. Conf. on Artificial Neural Networks*, pages 97–102, 1999.

[40] P.J. Huber. *Robust Statistics*. John Wiley and Sons, New York, 1981.

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

[42] C. E. Jacobs, A. Finkelstein, and D. H. Salesin. Fast multiresolution image querying. In *Computer Graphics Proc., Annual Conf. Series (SIGGRAPH 95)*, pages 277–286, 1995.

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

[44] M. Jones, P. Sinha, T. A. Poggio, and T. Vetter. Top-down learning of low-level vision tasks. *Current Biology*, 7:991–994, 1997.

[45] N. Kanwisher, J. McDermott, and M. M. Chunn. The fusiform face area: A module in humane extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17:4302–4311, 1997.

[46] A. Kleinschmidt, C. Buchel, C. Hutton, K. J. Friston, and R. S. J. Frackowiak. The neural structures expressing perceptual hysteresis in visual letter recognition. *Neuron*, 34:659–666, 2002.

[47] E. Kobatake and K. Tanaka. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiology*, 71:856–857, 1994.

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

[49] S. M. Kosslyn. *Image and Brain*. MIT Press, Cambridge, MA, 1994.

[50] Z. Kourtzi and N. Kanwisher. Representation of perceived object shape by the human lateral occipital complex. *Science*, 293:1506–1509, 2001.

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

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

[53] K. Linkenkaer-Hansen, J. M. Palva, M. Sams, J. Hietanen, H. J. Aronen, and R. J. Ilmoniemi. Face-selective processing in human extrastriate cortex around 120ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters*, 253:147–150, 1998.

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

[55] J. Liu, M. Higuchi, A. Marantz, and N. Kanwisher. The selectivity of the occipitotemporal M170 for faces. *Neuroreport*, 11:337–341, 2000.

[56] N. J. Majaj, D. G. Pelli, P. Kurshan, and M. Palomares. The role of spatial frequency channels in letter identification. *Vision Research*, 42:1165–1184, 2002.

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

[58] A. Martin, C. L. Wiggs, L. G. Ungerleider, and J. V. Haxby. Neural correlates of category specific knowledge. *Nature*, 379:649–652, 1996.

[59] M. F. Mendez and M. Ghajarnia. Agnosia for familiar faces and odors in a patient with right temporal lobe dysfunction. *Neurology*, 57:519–521, 2001.

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

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

[62] R. Nasanen. Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, 39:3824–3833, 1999.

[63] W. T. Newsome and E. B. Pare. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neuroscience*, 8:2201–2211, 1988.

[64] T. B. Penney, B. Maess, N. Busch, J. Derrfuss, and A. Mecklinger. Cortical activity reduction with stimulus repetition: A whole-head MEG analysis. *Cognitive Brain Research*, 16:226-231, 2003.

[65] D. I. Perrett, J. K. Hietanen, M. W. Oram, and P. J. Benson. Organization and function of cells responsive to faces in the temporal cortex. *Trans. Royal Soc. London*, B225:23–30, 1992.

[66] T. Poggio. On optimal nonlinear associative recall. *Biological Cybernetics*, 19:201–209, 1975.

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

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

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

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

[71] G. Rainer and E. K. Miller. Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, 27:179–189, 2000.

[72] M. D. Rugg and M. C. Doyle. Event-related potentials and stimulus repetition in direct and indirect tests of memory. In *Cognitive Electrophysiology*, pages 124–148, Boston, 1994. Birkhäuser.

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

[74] J. Sadr and P. Sinha. Exploring object perception with Random Image Structure Evolution. Technical Report AI-Memo-2001-06, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, March 2001.

[75] J. Sadr and P. Sinha. Random Image Structure Evolution (RISE). *Vision Sciences Society Abstracts*, 1:83, 2001.

[76] J. Sadr and P. Sinha. Characterizing object-specific neural correlates of perception. *Vision Sciences Society Abstracts*, 3:146, 2003.

[77] J. Sadr and P. Sinha. Exploring the neural correlates of object and face perception using magnetoencephalography and Random Image Structure Evolution. *Manuscript in preparation (a.k.a. Chapter 3)*, 2003.

[78] J. Sadr and P. Sinha. Object recognition and random image structure evolution. *Cognitive Science*, in press.

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

[80] P. G. Schyns and F. Gosselin. A natural bias for basic-level object categorizations. *Vision Sciences Society Abstracts*, 2:144, 2002.

[81] M. Seeck, C. M. Michel, N. Mainwaring, R. Cosgrove, H. Blume, J. Ives, T. Landis, and D. L. Schomer. Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport*, 8:2749–2754, 1997.

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

[83] C. R. Shelton. 3D correspondence. Master's thesis, MIT Dept. of Elect. Eng. and Comp. Sci., 1998.

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

[85] P. Sinha and T. A. Poggio. Role of learning in three-dimensional form perception. *Nature*, 384:460–463, 1996.

[86] P. Sinha and T. A. Poggio. High-level learning of early visual tasks. In *Perceptual Learning*, pages 273–298, Cambridge, MA, 2002. MIT Press.

[87] J. G. Snodgrass and J. Corwin. Perceptual identification thresholds for 150 fragmented pictures from the Snodgrass and Vanderwart picture set. *Perceptual and Motor Skills*, 67:3–36, 1988.

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

[89] J. G. Snodgrass and M. Vanderwart. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *J. Exp. Psych.: Human Perc. Perf.*, 6:175–215, 1980.

[90] K. Thoresz and P. Sinha. Qualitative representations for recognition. *Vision Sciences Society Abstracts*, 1:81, 2001.

[91] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-posed Problems.* W. H. Winston, Washington, D.C., 1977.

[92] G. Wahba. *Spline Models for Observational Data.* Series in Applied Mathematics, Vol 59, SIAM, Philadelphia, 1990.

[93] E. K. Warrington. Neuropsychological studies of object recognition. *Philosophical Transactions of the Royal Society of London*, B298:15–33, 1982.

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

[95] F. W. Young and C. H. Null. MDS of nominal data: The recovery of metric information with ALSCAL. *Psychometika*, 53(3):367–379, 1978.

[96] S. Zeki, J. D. G. Watson, C. J. Lueck, K. J. Friston, C. Kemard, and R. S. J. Frackowiak. A direct demonstration of functional specialization in human visual cortex. *J. Neuroscience*, 11:641–649, 1991.

I wonder about the trees.
Why do we wish to bear
forever the noise of these
more than another noise
so close to our dwelling place?
We suffer them by the day
till we lose all measure of pace,
and fixity in our joys,
and acquire a listening air.
They are that that talks of going
but never gets away;
and that talks no less for knowing,
as it grows wiser and older,
that now it means to stay.
My feet tug at the floor,
and my head sways to my shoulder
sometimes when I watch trees sway,
from the window or the door.
I shall set forth for somewhere,
I shall make the reckless choice
some day when they are in voice
and tossing so as to scare
the white clouds over them on.
I shall have less to say,
but I shall be gone.

- Robert Frost -