Title
Understanding how Cyclopean Perception Arises from Binocular Vision

Author
Guan, Phillip

Publication Date
2015

Peer reviewed|Thesis/dissertation

Understanding how Cyclopean Perception Arises from Binocular Vision

By

Phillip Guan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy

with the University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Martin S. Banks, Chair
Professor Christoph E. Schreiner
Professor Clifton M. Schor

Summer 2015

Abstract

Understanding how Cyclopean Perception Arises from Binocular Vision

by

Phillip Guan

Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor Martin S. Banks, Chair

In vision sensation and perception are rarely equivalent. Most notably, humans perceive the world with three-dimensional depth in a manner consistent with having a single, cyclopean eye centered above the nose even though the inputs to the visual system are a pair of binocular inputs each of which is two-dimensional. In this dissertation we address three aspects of cyclopean perception and how each is influenced by its underlying binocular inputs. In Chapter 1 we study how viewing distance affects retinal disparity and how the visual system is able to maintain a depth percept that is invariant to these changes. In Chapter 2 we examine how binocular disparity and monocular luminance cues are integrated to perceive edges. Finally, in Chapter 3 we consider how monocular images from the left and right eyes are combined to optimize the cyclopean image. Together these studies provide insights into how the visual system interprets retinal inputs and transforms them into a single perception of the world.

Acknowledgements

I moved from Seattle to Berkeley to start graduate school in 2009. At the time I was young, naïve, and the concept of finishing school and receiving a PhD just seemed like a neat idea. Now, almost exactly six years later, I'm returning back to Seattle a little bit less young, maybe a little bit less naïve, and the memory of watching my family drive away from my first Berkeley apartment seems like a lifetime ago. There are too many people that I'd like to thank for helping me realize that there is nothing in vision science or life so hopelessly complex and overwhelming that can't be addressed by framing the problem in the proper context, using the right tools, and working together with close friends.

The Banks Lab

I want to thank my adviser Marty for his insights and guidance through the past five years I've been his student. Despite my best efforts, your methodical approach to thinking about scientific questions, addressing looming problems head on, and communicating esoteric ideas is finally starting to wear off on me. You've helped me become a better scientist and individual and you will have a lasting influence on my career and life philosophies.

Peter you are an invaluable resource for getting things done in lab and interfacing with Berkeley. Without you the wheels would fall right off the car because Marty would be responsible for keeping things organized and we all know how big of a disaster that would be.

My colleagues in the Banks lab all of whom I also consider my friends. We've all experienced the neck stiffness of a bite bar, the eyestrain of a vergence-accommodation conflict, and the agony of a sign error in MATLAB. This means we have a special bond that can only be formed through shared hardship. David Hoffman, when I rotated in Marty's lab my biggest concern was that as a psychology the lab wouldn't be rigorous enough; working with you and seeing your technical acumen alleviated all of those concerns and helped convince me to join the lab. David Kane we first worked together at a point in my career where I knew almost nothing about psychophysics and vision, thanks for teaching me by example and getting our names out there on a great paper. Jamie I wish we had overlapped a little bit more because I had so many CS questions over the years; thanks for helping out with BASH and other random things that still make no sense to me. Robin thanks to you I am about to start at a dream job, I'm really sorry you didn't get that referral bonus because without you I am working at Apple; I can't wait to start working with you soon. Emily your enthusiasm for science is infectious and you always asked the best questions during lab meetings; I am hopeful that you can be a shining beacon of vision research amidst a sea of face perception. Bill you've worked on some really exciting projects and I appreciate being your lab neighbor and being able to bounce ideas off of you. Joohwan I'm also sorry we never got to publish together, but seeing how hard you worked was always an inspiration to me. You're going to make a great father soon. Paul I could always count on you to vote for the place with the biggest lunches, on a scientific note, your presentations had the best graphics and were always well-organized; both things I hope I'll be able to replicate in the future. Val I appreciate all of your statistics help and I'm really glad we've gotten to know each other better since you've left lab. Andy working together with you on the SFN poster

was a really good experience for me and I learned a lot (I hope you did too!); I admire how you had the fortitude to decide that all the authority figures in your life were wrong, and I hope I can do the same if the need ever arises. Last, but not least Marina. From our shared panic in the job search process, struggles with the BF display, and day-to-day science frustrations I don't think I could have made it through this past year with my sanity intact without being able to vent to you. You've been an incredible help for my research-related questions, but even better friend. I'm really looking forward to getting burgers and fries from Dick's and enjoying them with you and Pierre on your really expensive, but cheap-looking furniture in Washington.

Past Mentors

As an undergraduate at UW I was fortunate to have many opportunities that helped bring me to a place like Berkeley. Chris Neils introduced me to the signal processing concepts that lead me to medical imaging and eventually to vision science. His enthusiasm and encouragement helped spark my curiosity and interest in concepts that might have otherwise seemed opaque. Qiuming Yu and Dong Qin were my first two undergraduate research advisers. As a sophomore they trusted me enough to give me the freedom to run my own experiments and work on equipment that, if I broke, would disrupt the research of labs throughout the entire UW campus. Unfortunately Dong left shortly after I started at NTUF, but I will always remember how kind and genuine she was to me even though I was just an inexperienced undergrad. Qiuming trusted me enough to put me in a position to succeed as an independent research; I didn't realize it at the time, but the project I worked on was quite novel and has a high likelihood of remaining my most cited work. Even when I expressed a desire to change research directions Qiuming was supportive and helpful. I cannot express enough how grateful I am to both Dong and Qiuming for that first research opportunity. Xingde Li was my senior thesis adviser to whom I am also greatly thankful. In his lab I was given a project equal in scope to something a graduate student would have worked on and I was able to see experience firsthand my first research lab. I received several awards thanks to Xingde's mentorship and my experiences in his lab played a pivotal role in my acceptance into various graduate programs.

Friends

My friends, of which there are too many to name. I've been inspired and encouraged by all of you to try and accomplish so many things that I never would have even though about attempting without you guys. Some highlights include making three awesome music videos, choreographing an interpretive tree dance, lecturing at Berkeley, and interning at Imagineering. I cannot thank all of you enough for making me laugh, helping me re-examine and reshape my beliefs, and inspiring me become a better person.

There are a few individuals I am especially grateful for, first, my friends from college Neema, Jordan and Wayne. We've gone our separate ways since we graduated, but I hope someday we all end up in the place again so we can play NBA Live one more time (but only if Biedrins can win MVP). Being able to unwind with you guys over Yunnie, hoping for good DoBs at QFC, and doing CrossFit until we felt sick are boring, everyday memories that I will always cherish.

I would also like to thank Arunan who has been an amazing friend since our first days of at Berkeley. During the interview I felt like you and I might get along pretty well, but I wouldn't have guessed that we would end up developing the strong friendship that we have today. You've been an amazing lifting/science/tennis/dining-out/roommate/GoT buddy and I'm going to miss our frequent conversations about the serious and not-so-serious things in life (like celebrating a toasted or brown-cinnamon bagel).

Family

Jonathan and Crystal you guys were both still kids when I moved away. A lot has happened over the last six years and more and more I see that both of you have grown up into adults. I have been so impressed to see how mature both of you can be and I can't wait to see what the future holds in store for both of you. To my dad, a lot of the person I am today is thanks to you. Caring about issues like the environment, inequality, and doing things the right way, all of that I learned from your example. I hope things get better between everybody over the upcoming years. Finally I want to thank my mom, the most selfless person I know. It's only recently that I've been able to appreciate how much you care about Jonathan, Crystal, and me and how much you've done for us. I can't imagine how frustrating and thankless raising a single teenage is like much less going through that three times. When I was younger I was easily irritated and quick to anger, but I looked to your calm demeanor as a model for approaching conflict and I consider that one of my best qualities. I hope that someday, if I become a parent that I can be as kind and loving as you have been and continue to be to me.

Table of Contents

Chapter 1: Maintaining Depth Constancy from Disparity across Viewing Distance and Spatial Frequency

Introduction

The images on our retinas are constantly altered by changes in luminance, viewing distance, and perspective; but our perception of the world is often invariant to these effects. For example, when a ball is tossed at you the size of the ball on your retina increases as it travels towards you and the retinal disparity generated by that ball changes as well. Despite these changes both the size and apparent depth of the ball remain constant as the ball travels towards you. The ability of the visual system to derive the same percept from retinal images that vary in response to the environment is called constancy and has been observed in many domains including color (Maloney and Wandell 1986), size (Emmert 1881), contrast (Georgeson and Sullivan 1975), and depth (Wallach and Zuckerman 1963). Understanding how the visual system transforms physical retinal signals into the world we perceive provides critical insights into how visual information is processed by the brain and provides clues for how to create synthetic imagery that is immersive and realistic.

Depth constancy is a particularly interesting phenomenon given the emergence of virtual and augmented reality systems that aim to use realistic depth cues to create immersive experiences and environments. There are many depth cues that can be used to achieve this goal and they can be broadly classified into triangulation cues (stereo, motion, focus), perspective cues (perspective projection, relative size, texture gradients), and light transport cues (occlusion, shading, aerial perspective). Many of these cues are ordinal depth cues and only provide information about the depth ordering of objects in the world. Other depth cues including all triangulation cues and some perspective cues are metric depth cues and can provide information about how far away objects are relative to the point of fixation. Combined with an estimate of the fixation distance an observer can use metric depth cues to determine absolute distances to different objects in the world.

A reliable estimate of viewing distance is especially important for triangulation cues because the retinal signals generated by these cues are ambiguous without knowledge of the fixation distance. For example, in stereopsis the visual system cannot use disparity magnitude alone to create the perception of depth. As illustrated in Figure 1 the same retinal disparity can be attributed to an object that is far from fixation at a large viewing distance or an object relatively close to fixation at a near viewing distance. The visual system can resolve this ambiguity by estimating fixation distance from a combination of accommodation, vergence, and vertical disparity and then scaling the perceived depth from disparity accordingly. This allows the visual system to achieve depth constancy from disparity over a wide range of viewing distances (Collett et al. 1991, Rogers and Bradshaw 1993, Watt et al. 2005).

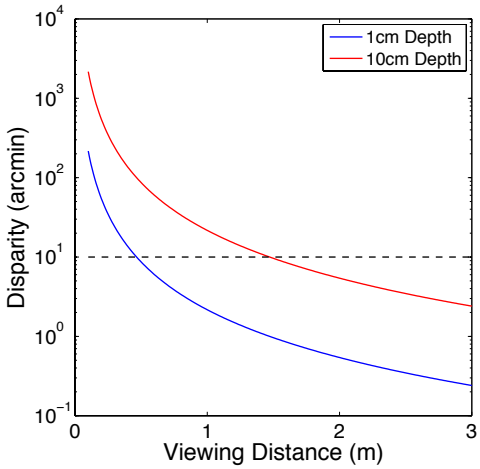


Figure 1 Disparity generated by objects 1cm and 10cm from fixation at distances varying from 10cm to 3m for an individual with an inter-pupillary distance of 63mm. The dashed line represents retinal disparity of 10arcmin and shows that the objects 1cm and 10cm from fixation can both generate the same disparities at different viewing distances. Thus 10arcmin of disparity associated with a viewing distance of 0.5m should generate a very different amount of perceived depth as 10arcmin at 1.5m viewing distance.

In addition to viewing distance estimation another issue that must be addressed in order to maintain depth constancy from stereopsis is the visual system's variable sensitivity to disparities across spatial frequency. In Figure 2 (from Kane et al. 2014) the perceived depth from disparity is visualized across multiple spatial frequencies and disparity magnitudes. There is a limited range of disparities and spatial frequency combinations that support depth perception; furthermore, the smallest and largest perceivable disparities vary with spatial frequency. The disparity sensitivity function is bandpass (roughly C-shaped on the left of the Figure 2) with peak sensitivity to disparity near 0.3cpd and decreasing sensitivity at higher and lower spatial-frequencies. A similar phenomenon is observed in contrast sensitivity (Campbell and Robson 1968). In luminance the optics of the eye act as a lowpass filter (Santamaría et al. 1987, Liang et al. 1994). This means that if a 10cpd and 50cpd grating have equal contrast in the world, the retinal contrast of the 10cpd grating will be much higher compared to the 50cpd grating. At suprathreshold contrasts the visual system compensates for this differential filtering and both gratings are perceived to have the same contrast even though the physical retinal contrasts are different (Georgeson and Sullivan 1975).

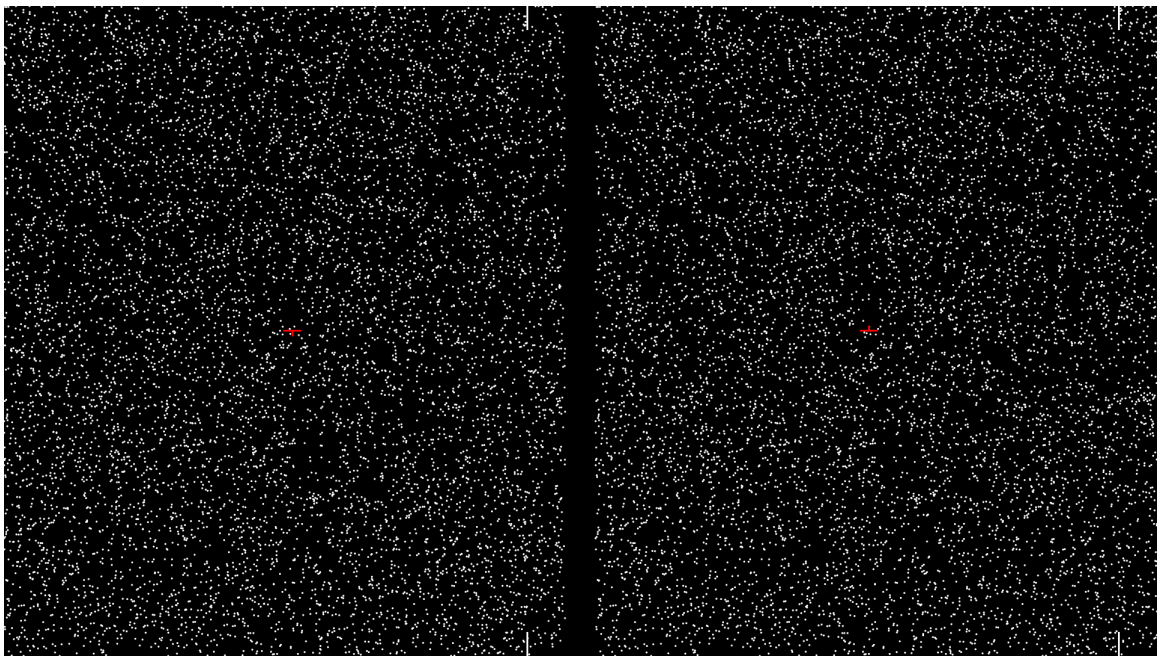


Figure 2 Cross fuse to view demonstration of smallest and largest visible disparities as a function of spatial frequency. Disparity increases from the left to the right of each monocular image and spatial frequency increases going from the bottom to the top of the monocular images. The smallest disparity supports depth (sensitivity limit) is not the same for all spatial frequencies; the largest disparities that support depth perception are also dependent up on spatial frequency (upper limit).

Perhaps the bandpass sensitivity in disparity perception results in depth constancy behavior similar to contrast constancy observed in luminance. However, there are differences between luminance and disparity processing that cause differences between depth constancy and contrast constancy. Unlike lowpass optical filtering in the luminance domain, the high spatial frequency attenuation observed in disparity sensitivity is a byproduct of the neural mechanism used to calculate disparity (Fleet et al. 1996, Qian and Zhu 1997, Nienborg et al. 2004). One key difference between luminance and disparity perception that arises because of this neural mechanism is the upper disparity-gradient limit (Tyler 1974).

The upper-disparity gradient limit describes the largest disparities that can sustain the perception of depth and it varies with spatial frequency. It is a consequence of determining retinal disparity via local cross-correlation (Banks et al. 2004, Filippini and Banks 2009, Kane et al. 2014). No analog exists in the luminance domain; in luminance increasing contrast will never reduce the perceptibility. Given the bandpass sensitivity in luminance and disparity, but noting the underlying physiological differences in luminance sensitivity and disparity sensitivity (optical vs. neural) we wondered if the visual system could compensate for apparent depth from disparity across various disparity spatial frequencies as observed in luminance spatial frequencies to achieve contrast constancy. We examined these questions in our first experiment by comparing the perceived depth of disparity gratings at different spatial frequencies to see if depth constancy was observed at various disparity magnitudes despite differences in the visual system's variable sensitivity to disparities at different spatial frequencies.

Experiment 1: Depth Constancy across Spatial Frequency

Observers. Four observers (two male and two female) 22-31 years of age participated in the study. One was an author and the other three were unaware of the experimental hypothesis.

Apparatus. The stimuli were presented on a mirror stereoscope with two CRT displays (Iiyama HM204DT). The lines of sight from the two eyes were reflected from mirrors near the eyes such that they were co-linear with a normal from the center of the respective CRTs. The experiment was conducted in a dark room, so the CRTs provided the only measurable light input to the eyes. The CRTs were set to a spatial resolution of 800x600 pixels. At the 115cm viewing distance, pixels subtended 1.5arcmin. By using anti-aliasing, we could create much smaller disparities. We estimate that the smallest displayable disparity was 2 seconds of arc; the smallest disparity we presented in the main experiment was 15 seconds of arc. Vergence distance was the same as the viewing distance. Refresh rate was 200Hz.

Stimuli & Procedure. Identical dynamic random-dot patterns were presented to the two eyes between trials. A fixation target composed of two binocular horizontal line segments and two dichoptic vertical line segments was also presented at all times. By monitoring the apparent alignment of the dichoptic segments, observers could make sure that fixation was accurate before initiating a stimulus presentation. They were instructed to maintain fixation on the fixation target during the stimulus presentation as well.

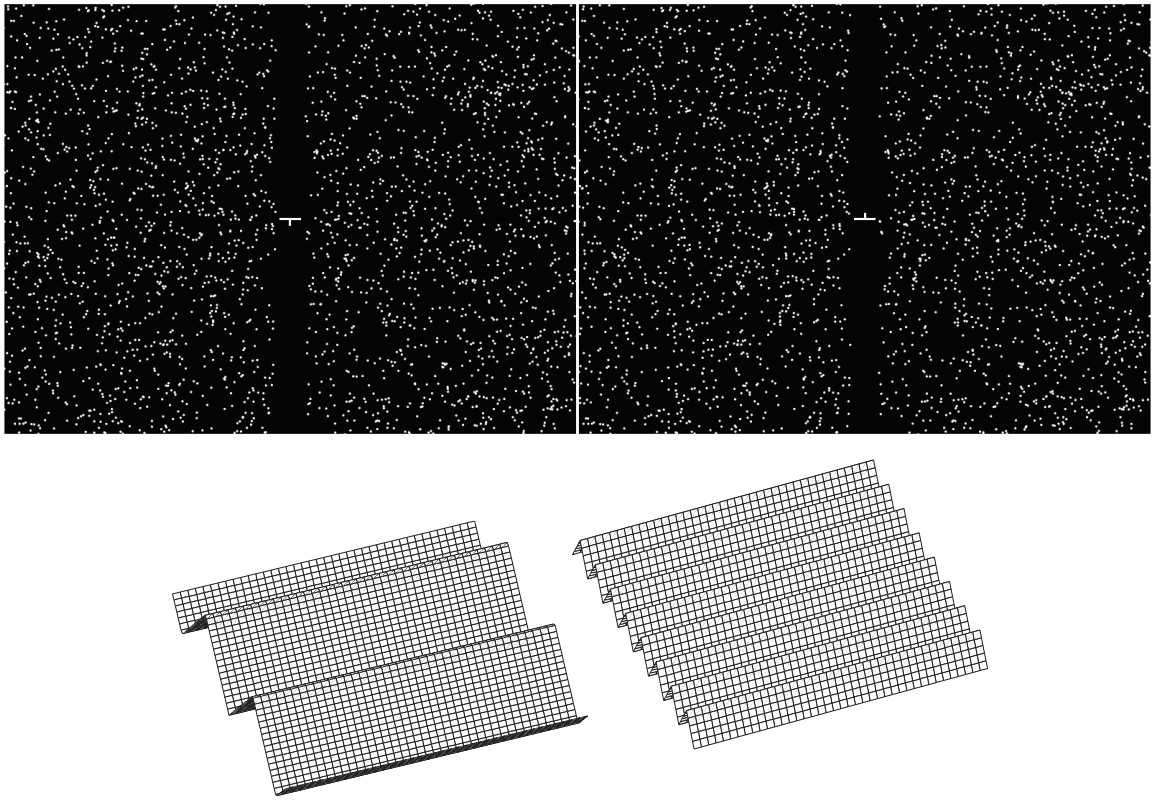


Figure 3 Example stimulus. In this example the 0.3cpd standard grating is shown to the left of fixation and a higher spatial frequency grating with less disparity is shown to the right of fixation.

The stimulus consisted of two triangular-wave gratings---a standard and a comparison---and is depicted in Figure 3. On each trial the standard and comparison gratings were randomly placed on opposite sides of a nonius cross and subjects were asked to identify the grating with greater apparent depth with a button press. The spatial frequency and disparity of the standard grating were fixed while the disparity and spatial frequency of the comparison grating were varied. In total there were eight standard gratings: each one had a spatial frequency of 0.3cpd with disparities of 0.25, 0.5, 1, 4, 16, 32, 64, or 96arcmin. The eight standard gratings had eight spatial frequencies (0.0625, 0.125, 0.25, 0.5, 0.75, 1, 1.5 and 2cpd) and the disparity of each comparison grating was varied using the method of constant stimuli yielding 64 psychometric functions. All conditions were randomly interleaved within an experimental session.

All stimuli were generated using PsychToolbox (Brainard 1997, Pelli 1997); the comparison and standard gratings were each 15° tall and 9° wide. The inner edges of the patches were 0.5° from the center of the fixation target. The dots in the stereograms were 3arcmin in diameter. Dots were refreshed at 200Hz. Dot density was 9 dots/deg^2 yielding a Nyquist frequency of 1.5cpd for each frame (Banks et al., 2004). However, new dots consistent with the simulated waveform were presented every 5msec, so the effective dot density (and therefore the effective Nyquist frequency) was much higher due to visual persistence (Lankheet & Lennie, 1996). Specifically, if the visual system integrated the information in *n* frames, the effective Nyquist frequency would be 1.5*n* (i.e., 6cpd for integration of four frames). There were unmatched dots (i.e., seen by one eye but not the other) near the left and right edges of the stimuli and the number of such dots increased with disparity amplitude. Thus the center of the signal stimulus looked like a coherent triangular wave while the edges appeared noisy. The central portion was large enough to make the depth judgement between standard and comparison easily. After each presentation, the observer indicated with a key press whether the waveform with the greater depth appeared on the left or right.

Data Analysis. Each psychometric function was fit with a cumulative Gaussian using psignifit (Fründ et al. 2011). Responses ranged from 0 to 100% where 0% indicated disparities where the comparison always appeared to have less depth than the standard stimulus and 100% corresponded to disparities at which the comparison always appeared to have more depth than the standard. The 50% point indicated the disparity of the comparison grating that appeared to have the same depth as the standard stimulus. Due to the disparity gradient limit the largest disparities that could support depth perception from higher spatial frequencies could not generate enough depth to match the perceived depth of the standard; these particular conditions could not be fit with a psychometric function. We discarded all psychometric functions with minimums greater than 25% and maximums less than 75% (see appendix). Results were similar across all observers so data was analyzed by combining each subject's individual trials. Approximately 9,000 trials were collected for three subjects and 11,000 trials were collected for one naïve subject. The group average is obtained by combining the trials for all subjects prior to fitting the psychometric functions.

Experiment 1: Results

In this first experiment the reference and standard are at the same viewing distance so we do not need to consider the effect of viewing distance on depth constancy for these results and instead we only consider how perceived depth from disparity is be affected by spatial frequency. Figure 4 shows three possible ways perceived depth from disparity might be affected by spatial frequency. The black lines in the panels represent the lower-disparity thresholds and upper-disparity limits from Kane et al. (2014). The u-shape of the lower thresholds indicates that a disparity grating at 0.3cpd is visible at a smaller disparity relative to other spatial frequencies. The eight colored lines in the panels represents the iso-perceived depth lines where the disparity at different spatial frequencies are perceived equally in depth compared to a 0.3cpd standard. If depth from disparity is only affected by the lower sensitivity thresholds (sensitivity rule) the iso-perceived depth lines will mirror the bandpass thresholds. If only the upper disparity limits affect perceived depth from disparity (upper limit rule) then more disparity at low spatial frequencies will be required to be perceived as the same depth relative to higher spatial frequencies and the iso-perceived depth lines will mirror the upper-limit. If spatial frequency has no effect on perceived depth from disparity (spatial frequency constancy rule) then the disparity required to create the same amount of depth sensation in the 0.3cpd standard grating and the iso-perceived depth line will be horizontal with zero slope.

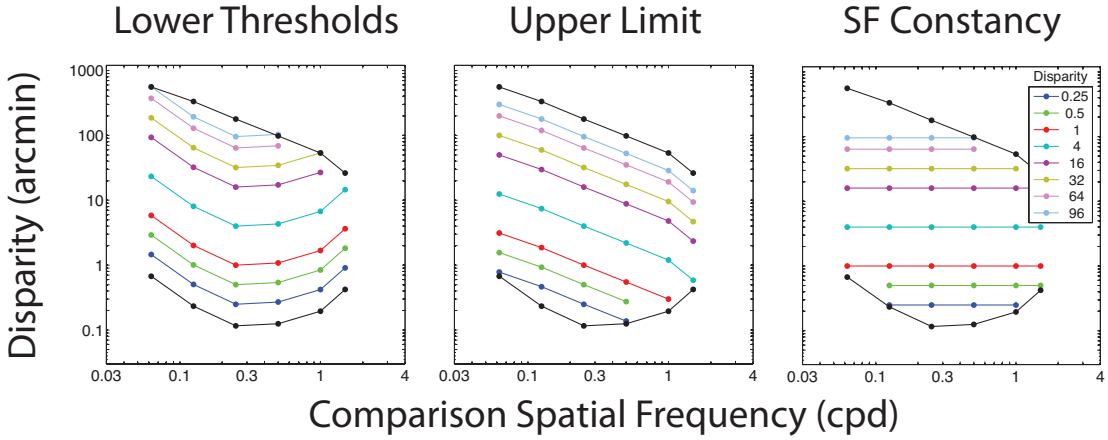


Figure 4 Predictions for perceived depth using the sensitivity rule, upper limit rule, and disparity constancy rule, solid black lines represent the lower disparity thresholds and upper disparity limits from Kane et al. (2014). For the sensitivity rule disparities closer to 0.3cpd are perceived to have the most depth and low and high spatial frequencies require more disparity to be perceived as having equal depth so the equal depth lines are bandpass in shape. In the upper limit rule higher spatial frequencies are perceived to have more depth than lower spatial frequencies so the equal depth lines have a negative slope. In disparity constancy spatial frequency has no effect on perceived depth so depth from disparity is perceived equally across all spatial frequencies and the equal depth lines are flat.

In the luminance domain Georgeson and Sullivan (1975) found that contrasts are perceived according to the sensitivity rule at low contrasts and according to the spatial frequency constancy rule at suprathreshold contrast. We might expect the same behavior for near-threshold and suprathreshold disaprities, but it is unclear how perceived depth from disparity might be affected near the upper limit. Figure 5 summarizes the individual and group results. In this plot the colored squares represent the eight 0.3cpd gratings that subjects used to match to the eight comparison gratings at spatial frequencies from

0.0625cpd to 2cpd. Subjects 2 and 4 could not reliably perceive disparities of 0.25arcmin so their data are excluded from that condition in the group average.

When disparities were small and near then disparity thresholds (<=1 arcmin) the iso-perceived depth lines mirror the sensitivity thresholds. When the disparity magnitude of the standard was 4arcmin and larger the iso-perceived depth lines were flat; i.e., perceived depth from disparity did not depend upon spatial frequency. At 16arcmins the comparison gratings at 1cpd and 2cpd approached the upper disparity limit and no amount of disparity at 1 and 2cpd resulted in as much perceived depth as the 16arcmin standard at 0.3cpd. At 32arcmin the depth-matching task could only be completed for spatial frequencies smaller than the standard spatial frequency of 0.3cpd, but depth constancy was still observed. When the disparity of the standard approached the upper-disparity limit at 64arcmin equal perceived depth occurred when the comparison's disparity was less than 64arcmin. This suggests that apparent depth from disparity declines precipitously as the disparity gradient-limit is exceeded rather than reaching a peak and gradually declining as values approach the upper limit.

In general the trends we observed in experiment 1 were similar to those observed by Georgeson and Sullivan for contrast constancy. Disparities near the lower thresholds are not perceived equally across spatial frequency, but once disparities reach a suprathreshold value they are. The major difference between depth constancy and contrast constancy is related to how disparities are perceived near the upper disparity limit. The shape of the upper disparity limit has no effect on the perceived depth from spatial frequency, but it does limit the largest disparities at which depth can be perceived and it affects higher spatial frequencies before lower ones.

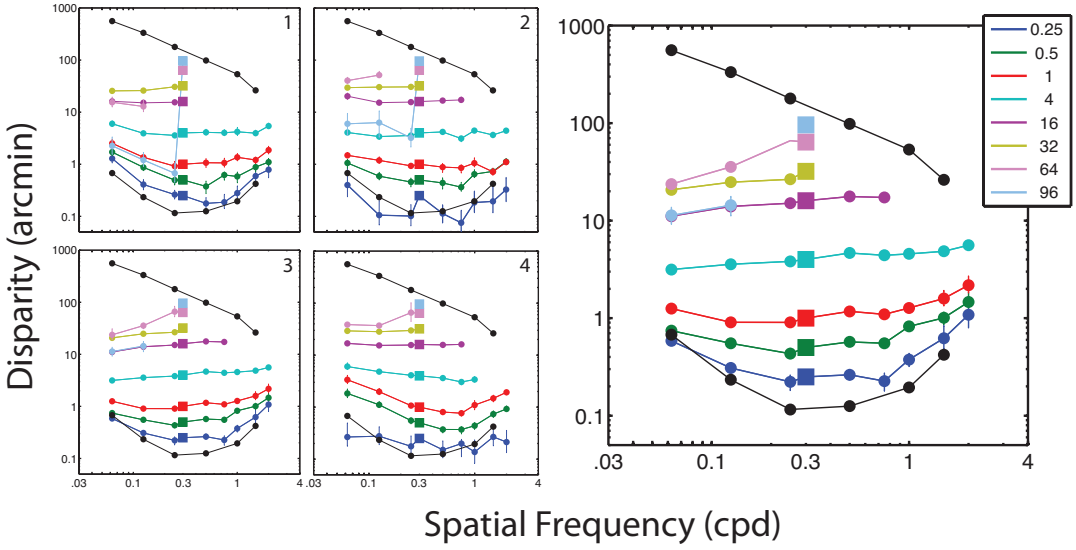


Figure 5 Perceived depth across corrugation frequencies. The disparity amplitude of the comparison stimulus that matched the perceived depth of the standard stimulus is plotted as a function of the spatial frequency of the corrugations in the comparison stimulus. The standard stimulus was a triangular-wave corrugation with a spatial frequency of 0.3cpd. It had one of eight disparity amplitudes as indicated by the colors and ordinate values of the squares. The comparison stimulus was also a triangular-wave, but at other corrugation frequencies. The disparity amplitude of the comparison that matched the perceived depth of the standard is indicated by the circular symbols. Colors indicate which standard stimulus was used in the match. The matching amplitudes were determined using the method of constant stimuli. Error

bars represent 95% confidence intervals. At 64arcmin spatial frequencies of 0.0625 and 0.125cpd are matched in perceived depth at disparities much smaller than 64arcmin while a 0.25cpd corrugation exhibits constancy. Three subjects were able to complete the task at 96arcmin, but a psychometric function could not be fit to the group data so we removed those results from the group average.

In summary there are three major disparity ranges of note in perceiving depth from disparity. First, at small disparities near the lower sensitivity threshold perceived depth is not constant and depends on spatial frequency. In the second region depth constancy is observed at suprathreshold disparities; i.e. the perceived depth from disparity is independent of spatial frequency. The third region encompasses disparities near the upper-disparity gradient limit. These three areas are illustrated in Figure 6. In this figure spatial frequency of each column and disparity of each row is fixed. The spatial frequency of each panel increases from left to right and the disparity of each row increases from top to bottom. Even though the physical disparity is the same for all panels the top row, the disparity magnitude is near threshold and the perceived depth is dependent of spatial frequency; here the central spatial frequency should appear to have more depth than the others. In the middle row the disparities are suprathreshold and all spatial frequencies should have the same apparent depth. In the last row the disparities are large and approaches the disparity gradient limit for the middle spatial frequency (and exceeds it for the largest spatial frequency); here the lowest spatial frequency appears to have the most depth and the highest spatial frequency should be a noisy without a coherent shape.

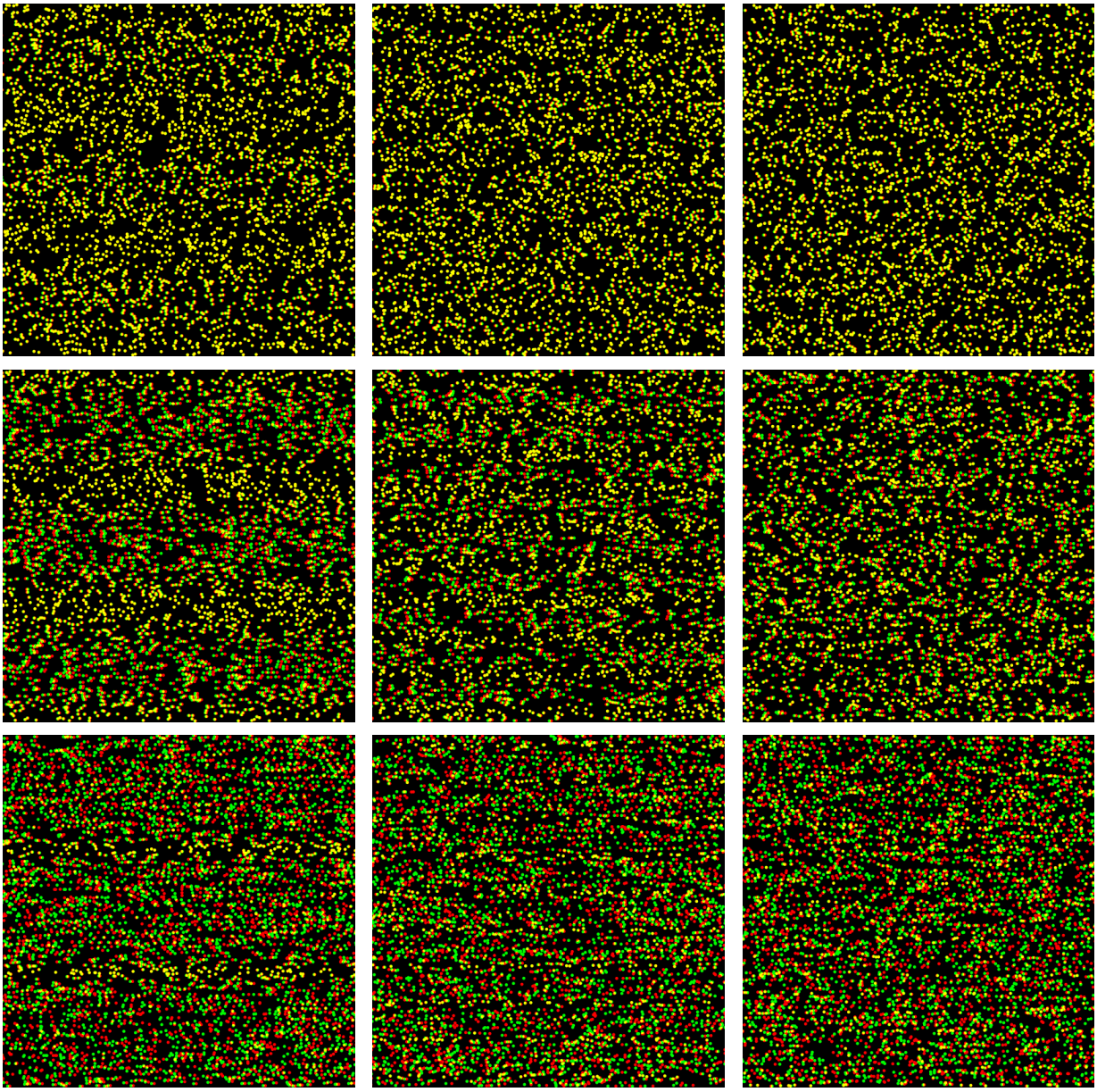


Figure 6 Perceived depth from disparity and spatial frequency. This figure should be viewed from approximately two feet away with red/green anaglyph glasses. The spatial frequency of each column and the disparity of each row is the same. Spatial frequency increases from left and right and disparity increases from top to bottom. Each one of the three rows roughly illustrates perceived depth from disparity near the disparity threshold, at suprathreshold disparities well below the disparity-gradient limit, and near the disparity-gradient limit respectively.

Figure 7 shows the disparity required in the comparison that was perceived as having the same depth as the reference for all eight spatial frequencies. The dashed line represents the spatial frequency constancy line where the same amount of disparity in the comparison is perceived as having the same depth as the standard. Spatial frequencies most similar to 0.3cpd, like 0.25cpd, are depth constant across all disparities. Below 1arcmin spatial frequencies much lower or higher than 0.3cpd required larger disparities to have equal perceived depth as the standard and these points are above the depth constancy line. This is consistent with the bandpass shape of the disparity sensitivity

function. Above 4arcmin and up to approximately 32arcmin all points coincide with the depth constancy line, here all spatial frequencies are perceived has having the same depth at the same disparity. Above 32 minutes of arc, the spatial frequency lines greater than 1cpd begin to terminate because they reach the disparity gradient limit first and no matches can be made to the standard. As the standard approaches the disparity gradient limit the lower spatial frequencies fall below the depth constancy line. This means that the perceived depth near the upper disparity gradient limit drops off very quickly rather than slowly (Kane et al. 2014). This behavior shows that increasing disparity leads to increased sensation of depth until the upper disparity limit is exceeded and the sensation of depth disappears suddenly.

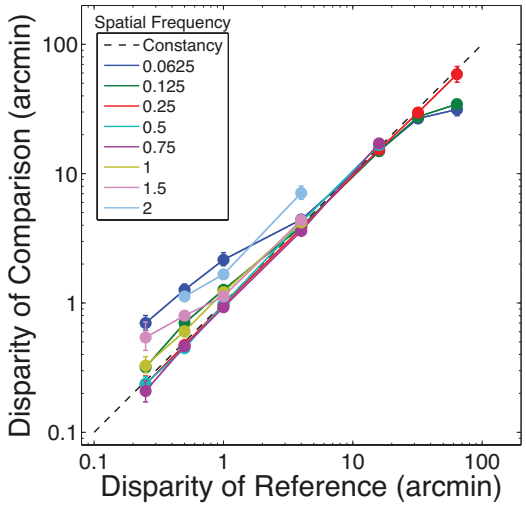


Figure 7 The disparity of the standard is shown on the abscissa and the disparity of the comparison at an equal perceived depth is shown on the ordinate. The dashed line along the diagonal shows where the data would lie for depth constant perception. Each colored line represents a different spatial frequency. Spatial frequencies that are not similar to 0.3cpd at disparities less than 1arcmin deviate from the depth constancy line, but by 4arcmin depth constancy is achieved for all spatial frequencies except 2cpd.

We examined the just-noticeable differences (JNDs) of the psychometric functions for each of the conditions in Figure 8; each panel shoes the same data, but plotted along different axes. Kane et al. (2014) found that JNDs at disparity threshold were larger than JNDs at the upper disparity limits and our results agree. The JNDs for each spatial frequency decreased as disparity increased, but increased again at the largest disparity of 64arcmin. The increase in JNDs at 64arcmin is likely due to inconsistently perceiving depth from the standard as it approaches its upper disparity limit. There appears to be a slight spatial frequency dependent effect on JND at smaller disparities with the smallest discrimination thresholds occurring at 0.3cpd and increasing JNDs as spatial frequency decreases and increases. As disparity magnitude increases the differences in JNDs across spatial frequencies appears to diminish although it is difficult to tell because higher spatial frequencies approach the upper limit and cannot be matched to the standard.

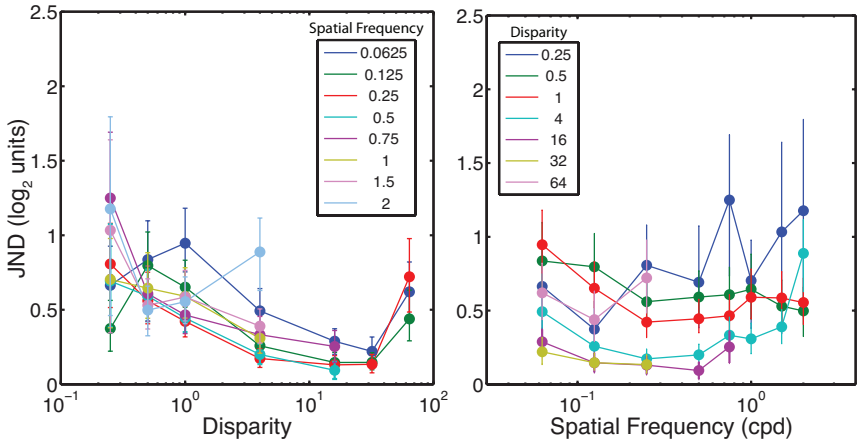


Figure 8 Slopes of the psychometric functions. In the left panel the absicca represents disparity of the standard and the ordinate is the slope of the fitted psychometric functions. Slope decreases for all spatial frequencies as disparity increases until the standard approaches the disparity gradient limit and the slopes once again increase. In the right panel the absicca is replaced by spatial frequency. Here it is clear that disparities well above threshold (4, 16, and 32arcmin) have the smallest JNDs.

Experiment 2: Depth Constancy across Spatial Frequency and Viewing Distance

In order to achieve depth constancy from disparity, the visual system must compensate for changes in disparity sensitivity to different spatial frequencies and disparity magnitude simultaneously. Like retinal disparity, the retinal spatial frequency of disparity signals changes with viewing distance, however, spatial frequency increases linerally with viewing distance while disparity decreases inversely with the square of viewing distance (Figure 9). We wanted to study both of these effects by repeating experiment 1 and placing the standard and reference gratings at different viewing distances (Figure 10). The visual system primarily uses accommodation, vergence, and vertical disparity as cues to estimate viewing distance (Collett et al. 1991, Rogers and Bradshaw 1993, Watt et al. 2005), in order to study the effects of viewing distance on depth constancy we had to ensure that those three cues were presented correctly. This could not be accomplished using traditional displays like the one used in the first experiment because such displays can only present stimuli at a fixed focal distance. This means that the vergence, vertical disparity, and accommodation can only be correct for a single viewing distance and presenting stimuli at any other distance generates a vergence-accommodation conflict and leads to incorrect viewing distance estimates (Watt et al. 2005, Hoffman et al. 2008). We addressed this by using a multi-focal plane display (Love et al. 2009), which allowed us to manipulate both the focal distance of our displays dynamically so that accommodation, vergence, and vertical disparities cues in our stimuli could be manipulated correctly for stimuli presented at different viewing distances.

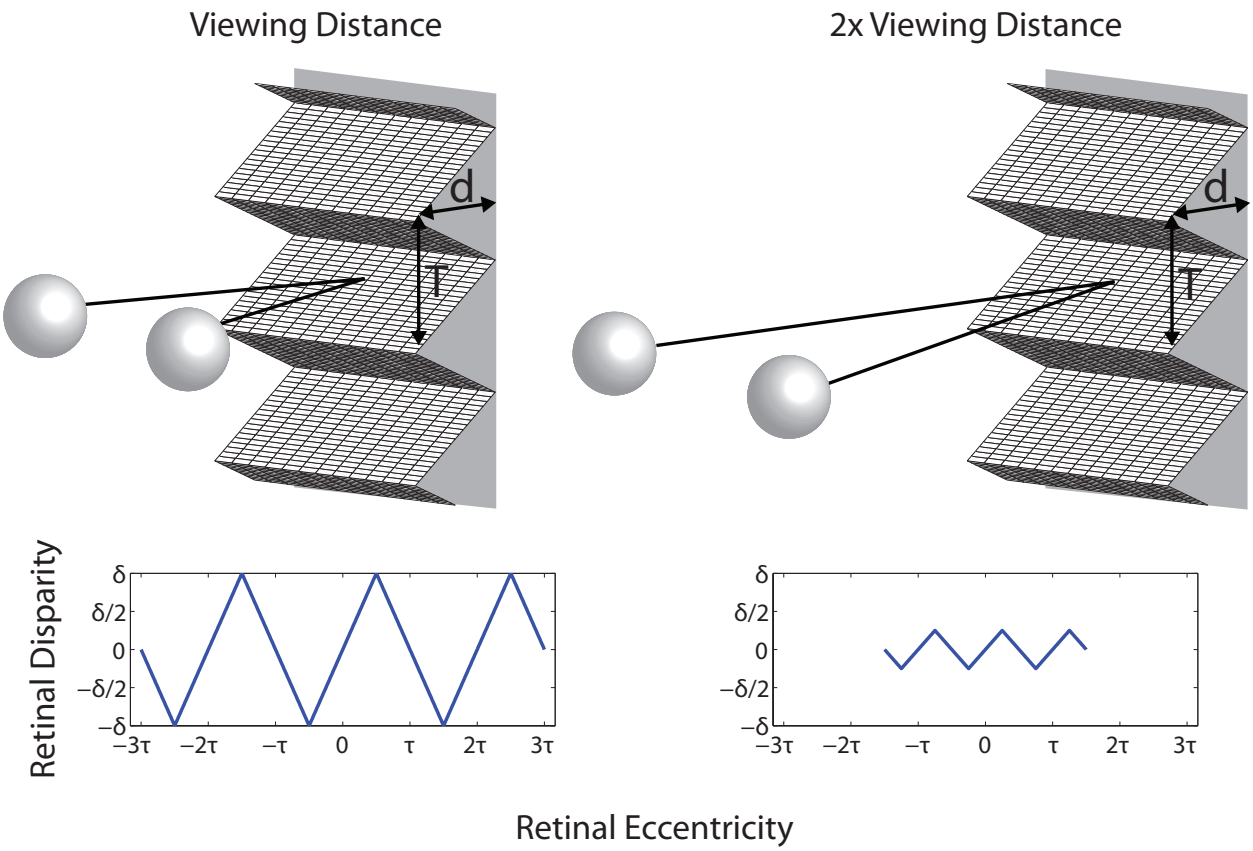


Figure 9 Here we show how retinal disparity and spatial frequency change for the same triangular wave grating viewed at two different fixation distances. Increasing the viewing distance by 2x reduces disparity by from δ to 0.25*δ and halves the period from 2τ to τ. The variables d and T represent the metric peak-to-trough depth and period of a grating in the real world. The variables δ and τ represent the disparity and period of the grating on the retina. In our experiment we manipulated d (1cm, 4cm, or 16cm) and 1/τ (0.125cpd, 0.3cpd, 0.5cpd, 1cpd) to study viewing distance effects on depth constancy.

Observers. A different set of observers was used compared to experiment 1, one was male and the other three were female, ages 20 to 27. One was an author and the remaining subjects were naïve to the purpose of the experiment; all had normal stereoacuity as measured by a Randot Stereo Test. In this experiment, stimuli were rendered using perspective projection which means that each observer was presented different depending on their inter-pupillary distance (IPD). Our subjects had measured IPDs of 59mm, 61mm, 61mm, and 63mm. In all of our analyses, we used an inter-pupillary spacing of 61mm.

Apparatus. More details of the multi-focal plane display can be found in Love et al. (2009). Briefly, this display allows us to manipulate the vergence distance and accommodation distance of our stimuli. In this experiment vergence and accommodation distances were always equal and set to either 77cm (1.3D) or 143cm (0.7D) from the subject., the haploscope display contains two birefringent lenses for each eye that are synchronized to the refresh rate of the monitors (180hz). Each eye has two lenses in series and each lens has two focal states, thus each eye is presented with four focal planes each refreshed at 45Hz. We calibrated the device so that when the eyes were at a vergence angle

corresponding to 1.3D the optic axis of the left and right eyes would be orthogonal to each eye's arm of the haploscope.

Stimulus and Procedure. We varied the physical peak-to-trough depth in centimeters of a triangular wave comparison stimulus until it matched the apparent depth of a standard with a fixed peak-to-trough depth (1, 4, or 16cm; d in Figure 9) at a different distance. If the standard grating was rendered and displayed at the far distance of 0.7D, then the comparison grating was rendered and displayed at near distance of 1.3D, and vice versa (Figure 10). The spatial frequency of the standard was always 0.3cpd ($1/2\tau$ in Figure 9). The spatial frequency of the comparison gratings was 0.125, 0.3, 0.5, or 1cpd. There were 24 total conditions (three standard depths each presented at two viewing distances compared to four spatial frequencies).

These three peak-to-trough depths and four spatial frequencies were chosen to cover the range of behaviors we observed from the different spatial frequencies and disparities examined in experiment 1. The four spatial frequencies span a large range of the visible spatial frequencies without requiring a large field of view (0.125cpd spans two cycles for a 17° field of view) or high dot density (for the dot density used in our stimuli the Nyquist limit was 2.17cpd). The peak-to-trough depths result in disparities that the three areas of note in experiment 1: near the lower sensitivity limit, suprathreshold disparities that exhibit depth constancy for all spatial frequencies, and near the disparity-gradient limit.

In order to compare stimuli having different viewing distances, we had to present stimuli using a two-interval, forced-choice method outlined in Figure 10. In one interval we would present a stimulus at 0.7D and the second we would present a stimulus at 1.3D and the order in which they were presented was randomized. During each interval, we presented two images. The first image was a dichoptic nonius cross to ensure that subjects adjusted their vergence and accommodation distance to be consistent with the vergence and accommodation distance of the RDS stimulus. We gave subjects an unlimited amount of time to align the nonius cross and upon a button press would show a RDS triangle wave grating for one second. After the presentation of the RDS, the second interval would begin. This interval also presented subjects with a dichoptic nonius cross and RDS triangle-wave. After seeing both intervals, subjects were asked to identify the interval in which the RDS triangle-wave appeared to have greater peak-to-trough depth. We used the method of constant stimuli to vary the metric peak-to-trough depth of the comparison stimulus to find the point at which subjects picked the comparison and standard gratings equally often. A total of 100-150 trials were collected for each subject in each condition (excluding the conditions near the upper gradient limit) resulting in approximately 2500 total trials per subject. All conditions (all four spatial frequencies and both viewing distance standards) for one of the three reference stimulus depths (1, 4, or 16cm) were interleaved within sessions. Thus we had three different possible session types, each with eight interleaved spatial frequency and viewing distance conditions.

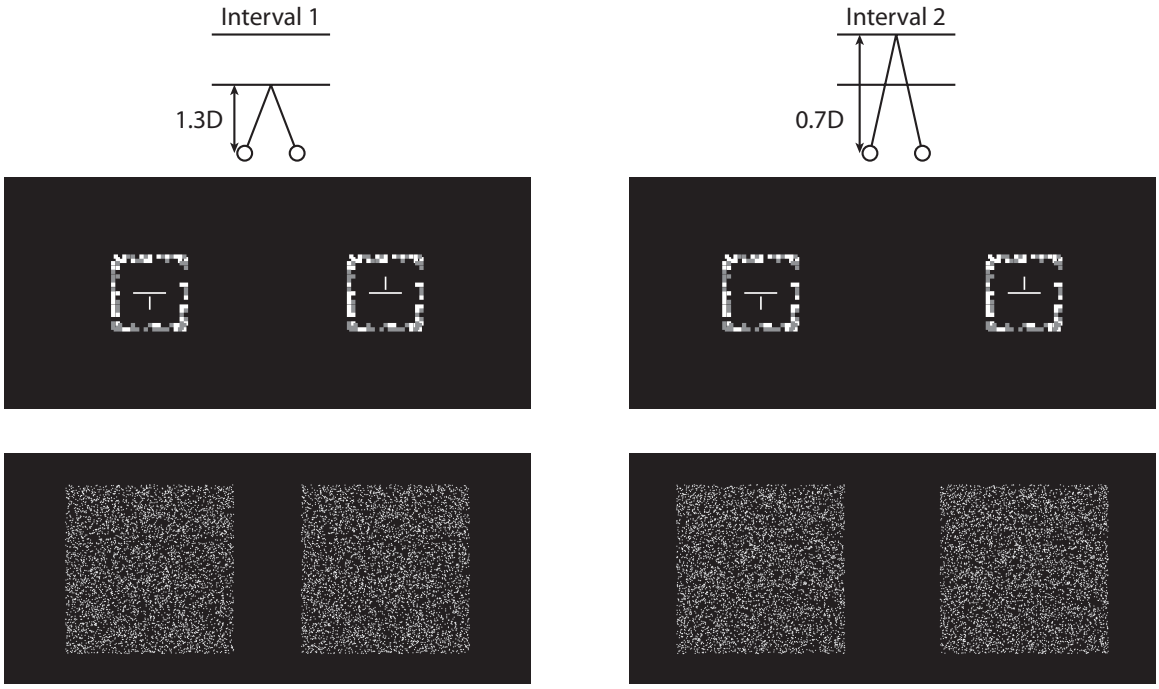


Figure 10 Two-interval, forced-choice stimulus presentation on multi-focal plane display. Each interval consisted of a fixation cross that was presented indefinitely at 0.7 or 1.3D to allow subjects to fixate and converge to the proper viewing distance. After a button press, a triangular RDS waveform was presented for one second at the same viewing distance as the nonius cross. The second interval was the same as the first except it was shown at the other viewing distance. Within each experimental session, the standard could be placed at either viewing distance or either interval.

Rendering Stimuli. We used perspective projection to ensure that the vertical disparity cue in our random dot stereograms (RDS) were consistent with their intended viewing distances. Figure 11 illustrates the viewing geometry of our display for a viewing distance of 1.3D and equations 1-4 were used to calculate the xy coordinates for our random dot stereograms (Woods et al. 1993, Held and Banks 2008). In order to use these equations we first defined the xyz coordinates of a triangular wave grating in 3D space; the phase of the standard and reference gratings was independently and both were set to one of nine possible values from 0° to 180° in increments of 30°. We calculated the left- and right-eye perspective projections for 7500 points over a 20°x20° area (18.75 dots/degree^2 which translates to a Nyquist limit of 2.17cpd, Banks et al. 2004). After computing the perspective projection dots at the edge of the stereogram were visibly translated depending on their z-value and the pattern of dots along the edges of the RDS could be used as a depth cue. We eliminated this cue by cropping the image to use just the central 17°x17° portion of the RDS (Figure 10).

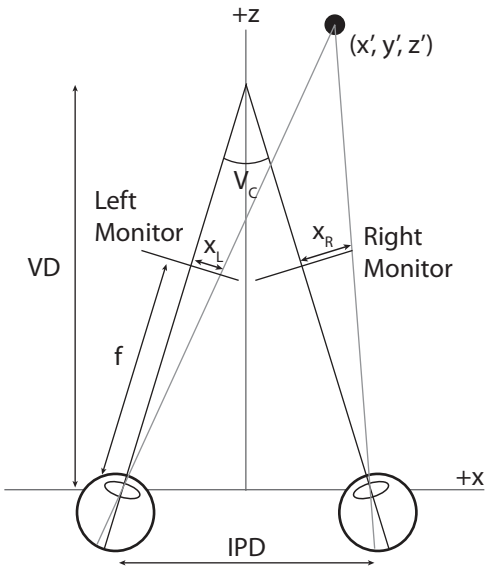


Figure 11 Multi-focal plane display rendering geometry. Our display was calibrated for a 1.3D viewing distance. Thus for stimuli presented at 0.7D and 1.9D we subjected each monitor to a virtual rotation to ensure that the line of sight from the nodal point of each eye remained normal to the surface of the monitors.

$$x_L = f\tan\left[\tan^{-1}\left(\frac{IPD/2 + x'}{z'}\right) - \frac{V_c}{2}\right] \quad \textbf{(1)}$$

$$x_R = f\tan\left[-\tan^{-1}\left(\frac{IPD/2 - x'}{z'}\right) + \frac{V_c}{2}\right] \quad \textbf{(2)}$$

$$y_L = \frac{fy'}{z'\cos(V_c/2) + (x' + IPD/2)\sin(V_C/2)} \quad \textbf{(3)}$$

$$y_R = \frac{fy'}{z'\cos(V_c/2) - (x' - IPD/2)\sin(V_C/2)} \quad \textbf{(4)}$$

Uniformly sampling 7,500 points on the surface of our waveforms and calculating the RDS images resulted in dot density variations within the monocular images, which is undesirable because changes in dot density act as a monocular depth cue. To ensure that the distribution of dots in our images was uniform we uniformly sampled a vertical plane located at the center of the waveform (either 0.7D or 1.3D). Then we projected the xyz points of the plane onto the surface of our triangular wave gratings to determine the xyz coordinates to use when calculating the perspective projections and the resulting left- and right-eye monocular images have uniform dot density.

The multi-focal plane display was calibrated for a viewing distance of 1.3D. A virtual rotation and translation is necessary to display stimuli at 0.7D correctly. We simulated vergence by translating the left- and right-eye images to correspond to the correct vergence angle (in Figure 11 if VD is increased then the optic axis of the eye intersects the

left and right monitors at different locations). When the vergence angle of the eyes changes the monitors are no long orthogonal to the optic axis of the eyes. To compensate for this change we virtually rotated the triangle-wave about the z-axis in opposite directions for each eye so that the perspective projections were correct for 0.7D.

Data Analysis. Fitting methods were unchanged from experiment 1. All conversions from metric units to disparities were calculated assuming an IPD of 61mm because our observers had IPDs of 59, 61, 61, and 63mm, therefore we used a 61mm inter-pupillary distance in our analysis. This assumption affects computed disparities presented in the results, but did not affect stimulus creation or presentation.

Experiment 2: Results

In experiment 1 depth from suprathreshold disparities were perceived as constant across spatial frequency (flat line in Figure 5). In experiment 2 we examined the effect of viewing distance and spatial frequency on depth constancy simultaneously, if spatial frequency has no effect on the perceived depth from disparity then the data should again lie on flat lines. However, the standard and the comparison gratings are not at the same vergence and accommodation distance so they should have different disparities when they are perceived to have equal depth. More specifically when the standard is farther than the comparison then the disparity of the comparison should be larger than the standard because disparity magnitude is inversely proportion to the square of viewing distance. This is illustrated in Figure 12, the colored squares in the right panel represent the disparity of the standard with different peak-to-trough depths and the dashed lines represent the corresponding predicted disparities for the comparisons that are depth constant with respect to spatial frequency (line is flat) and viewing distance (disparity of the comparison is larger than the disparity of the standard).

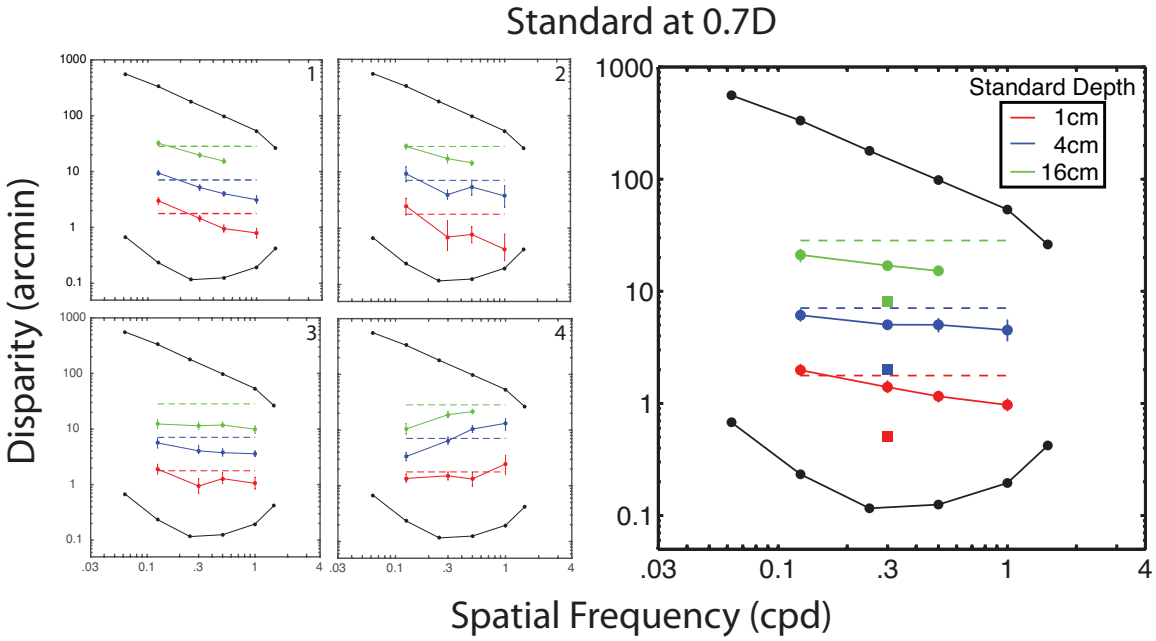


Figure 12 Perceived depth from disparity across spatial frequency and viewing distance. Here we show individual subject data and the combined group data when the reference grating is at 0.7D. In each panel

the abscissa represents the spatial frequency of the comparison grating and the ordinate represents disparity (calculated assuming 61mm IPD). The colored squares in the group averages show the disparity of the standard.

The group results in Figure 12 show that viewing distance is accounted for in our task (comparison iso-perceived depth lines are above the standard near the dashed lines) and that spatial frequency has a effect on perceived depth from disparity at the four spatial frequencies. The negative slope in the data indicate that less disparity is required to match the perceived depth of the reference at higher spatial frequencies, i.e. higher spatial frequencies are perceived to have more depth than lower spatial frequencies. However, only two of the four subjects exhibit this behavior. One subject had the opposite effect and perceived higher spatial frequencies as having less depth (positive slope) and the other showed relatively little differences in perceived depth from disparity over the four spatial frequencies (nearly zero slope).

Figure 13 corresponds to conditions where the standard is closer than the comparison. The disparities of the 1, 4, and 16cm peak-to-trough standards are again shown as boxes and the depth constancy predictions are shown as dashed lines. Here depth constancy line is below the disparity of the standard because the comparisons are farther away than the standards and should have less disparity in order to be perceived as having the same depth. The individual subjects again have different spatial frequency effects for perceiving depth from disparity and their biases remain unchanged from before. The iso-perceived disparity lines for the 4 and 16cm standard grating are very similar to the trends seen in Figure 12. Two subjects have a negative slope (higher spatial frequencies appear to have more depth) for two subjects, one has slopes nearly equal to zero (spatial frequency has no effect on perceived depth), and the last subject has a positive slope (higher spatial frequencies are perceived to have less depth). The 1cm standard grating is slightly different from before. Here the theoretical disparity that the comparison should be equal to approaches the lower sensitivity thresholds. Subjects 2 and 3, show the characteristic bandpass sensitivity that mirrors the sensitivity limit which was also observed in experiment 1. Subjects 1 and 4 do not appear to be affected by the smaller disparities because their iso-perceived depth lines appear similar in Figure 12 and Figure 13.

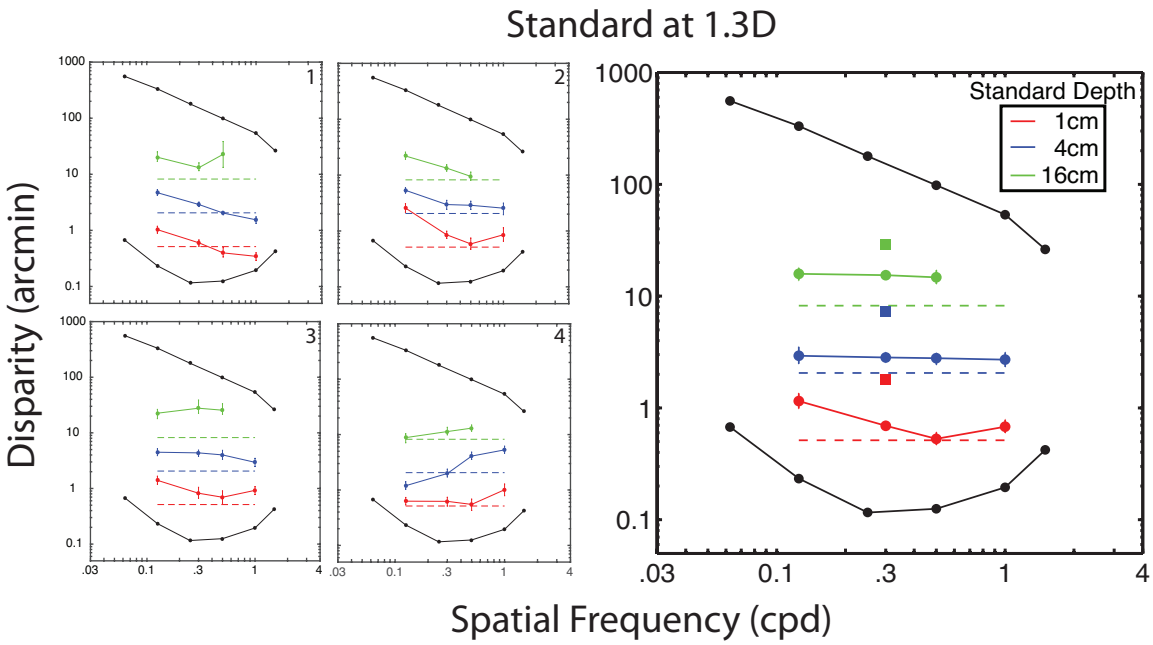


Figure 13 Perceived depth from disparity across spatial frequency and viewing distance. Here we show individual subject data and the combined group data when the reference grating is at 1.3D. In each panel the abscissa represents the spatial frequency of the comparison grating and the ordinate represents disparity (calculated assuming 61mm IPD). The colored squares in the group averages show the disparity of the standard.

Overall the combined group data in Figure 12 and Figure 13 are consistent with the behaviors observed in experiment 1. Disparities near the lower threshold appear to mirror the bandpass shape, suprathreshold disparities are nearly depth constant, and at extreme disparities near upper-disparity gradient limit depth from disparity is underestimated (requires more disparity than predicted). However, the single subject variability of responses to spatial frequency is quite different than experiment 1 where spatial frequency had no effect on perceived depth from suprathreshold disparity. One difference that might explain the differences in how individuals perceived depth as a function of spatial frequency in experiment 2 could be the perspective projection cues present in experiment 2, but not experiment 1. The RDS triangular waves in experiment 2 were rendered using perspective projection and the fused 3D shape contained these perspective cues that were not present in experiment 1(i.e. triangular sections at larger elevations appeared more skewed). These cues are more prevalent at higher spatial frequencies and different subjects could have different biases for how they integrate the perspective cue with the disparity cue. Using a two-interval forced-choice task instead of a two-alternative forced-choice task could further increase this bias among observers because the 2IFC task requires observers to remember a perceived depth and doesn't allow subjects to directly compare the standard and comparison waveforms.

We re-plot data from Figure 13 in Figure 14 to show how the data correspond to a depth constancy prediction; the ordinate shows the metric depth of the standard grating and the ordinate shows the metric depth of the comparison that appears to have the same depth of the standard at 1, 4, or 16cm. The dashed black line represents the depth constancy where the standard and comparison gratings that appear to have the same

perceived depth also have the same metric depth. If subjects are able to compensate for changes in viewing distance and spatial frequency, then all data will lie along the dashed line. From Figure 7 we expect a deviation from the line at least at 1cm due to the bandpass nature of disparity sensitivity. Prior studies on viewing distance-related disparity scaling (Collett et al. 1991, Rogers and Bradshaw 1993, Watt et al. 2005) suggest that the 4cm and 16cm standard gratings which generate suprathreshold disparities should fall along the depth constancy line.

When the reference grating is at 0.7D, the depth of the gratings at 1.3D is overestimated relative to the standard a grating with less physical depth in the comparison appears equal to the reference. The opposite is true when the reference is at 1.3D; here the depth from gratings at 0.7D is underestimated relative to the reference and a grating with more physical depth in the disparity appears equal to the standard. The error in the magnitude of the depth matching suggests that the subjects do not have a correct estimate of viewing distance, however, the error appears to be the same for all spatial frequencies. This means that the error in viewing distance compensation is not specific to any individual spatial frequencies. For both conditions, the magnitude of the error is largest for the smallest peak-to-trough depth which is in agreement with experiment 1.

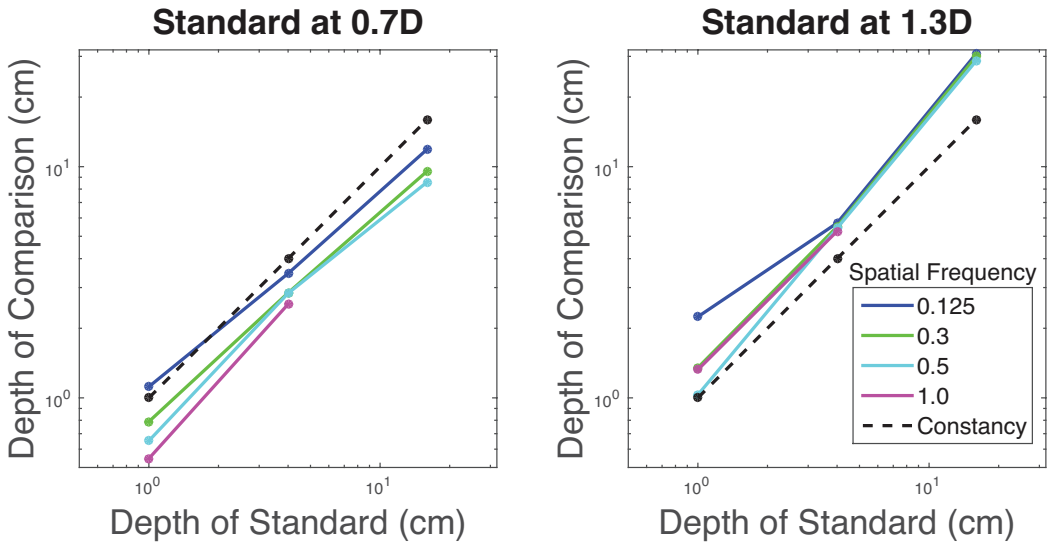


Figure 14 Depth constancy in metric depth for both viewing distance conditions. The depth of the standard is shown in the abscissa (1, 4, or 16cm) and the depth of the comparison that is perceived to match the standard in depth is shown on the ordinate. The dashed black line shows the depth constancy prediction when the visual system matches the same peak-to-trough depth in the standard to the same peak-to-trough depth in the comparison at a different viewing distance. Each of the four spatial frequencies are represented by a different color.

The effect of viewing distance can be visualized by converting the metric depth values from Figure 14. There are two different predictions highlighted in Figure 15. The dashed red line represents disparity matching, along this line the disparity of the standard and comparison are matched. Points along this line represent a failure of the visual to compensate for differences in viewing distance between the standard and comparison; the dashed black lines are the constant depth lines and represent the theoretical disparities of the comparison that would result in depth constancy. The depth constant line is above the

disparity matching line when the standard is at 0.7D because the comparison is closer than the reference so more disparity should result in the same perceived depth. Simiarly the depth constant line is below the disparity matching line when the standard is at 1.3D because the comparison is farther away and should generate less disparity to be perceived as having the same depth. We find that the disparities required to match the comparison to the standard are closer to the depth constant prediction than the disparity matching prediction. We can see how accurately viewing distance is estimated by seeing how closely the lines fall to the black depth constancy prediction. In both cases we can see that the magnitude of the viewing distance difference is slightly underestimated, but close to veridical. Figure 15C shows the estimated viewing distances given the disparities that appear to match the reference in depth.

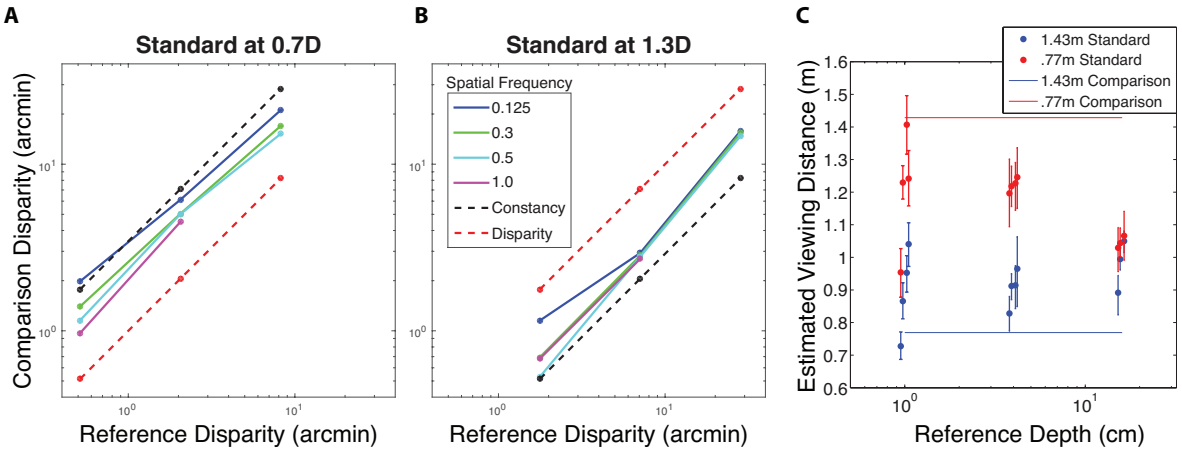


Figure 15 Depth constancy in minutes of arc for both viewing distance conditions. **A)** The disparity of the standard at 0.7D is shown in the abscissa and the disparity of the comparison that is perceived to match the standard in depth is shown on the ordinate. The dashed black line shows the depth constancy prediction when the visual system matches the same peak-to-trough depth in the standard to the same peak-to-trough depth in the comparison at a different viewing distance. The dashed red line shows the disparity matching prediction where no viewing distance compensation is made and depth matching is done using only the disparity magnitude. **B)** Same as A, except the standard is at 1.3D instead of 0.7D. Each of the four spatial frequencies is represented by a different color. **C)** Assuming disparity is perceived without error we calculated the visual system's estimated viewing distance for each condition. The blue line represents the viewing distance used to render the comparison when the standard is at 0.7D (shown by the blue dots) and the red line represents the viewing distance used to render the comparison when the standard is at 1.3D (shown by the red dots). There appears to be a systemic error in estimation of depth for all conditions. When the standard has a large depth at a near viewing distance (i.e. approaches the disparity gradient limit) depth is greatly misperceived in the comparison and the estimated viewing distance is not large enough.

We again examined the JNDs for each of our conditions; there are several notable differences in the results in experiment 2 compared to experiment 1. Experiment 2 was conducted using a two-interval, forced-choice procedure instead of a two-alternative forced-choice procedure so the magnitudes of the JNDs are higher. In experiment 1 disparity was the main predictor of slope with spatial frequency playing a more minor role, the opposite appears to be true in experiment 2. Here JND is consistent across disparities, but highly dependent upon spatial frequency. Somewhat puzzling is that there is no effect

of disparity on JND, it could be that the noise in remembering the two intervals is larger than any noise from the estimation of the depth itself.

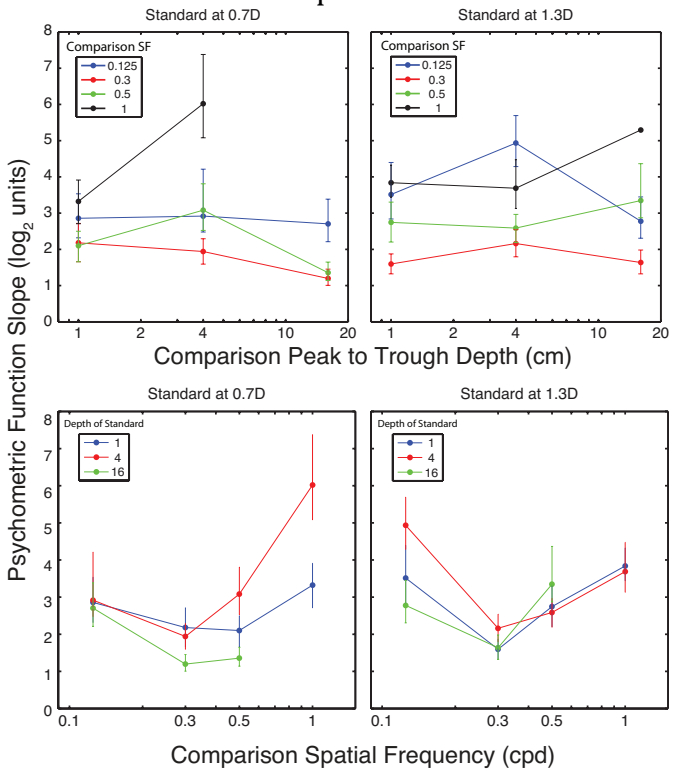


Figure 16 Top panels represent JNDs for different spatial frequencies. Absicca is peak to trough depth in metric units and ordinate is the slope of the psychometric function in log units. Bottom row represents JNDs for different peak to trough depths. Oridnate is again slope of the psychometric function in log units, but absicca is now spatial frequency of the comparison grating.

Experiment 3: Perceived Depth with Vergence-Accommodation Conflict

In natural viewing, the vergence and accommodation responses of the visual system are linked. Diverging the eyes means looking farther, which is accompanied by a change in dis-accommodation. Converging the eyes means looking at a nearer point and is accompanied by an accommodation to that location. In traditional stereo3D displays the visual system undergoes a vergence eye movement in response to disparities in stereo imagery, but maintains fixation on the display to keep the retinal image sharp resulting in a vergence-accommodation conflict. The vergence-accommodation conflict has been shown to result in eye fatigue and eyestrain (Hoffman et al. 2008, Shibata et al. 2011). Incorrect vergence and focal cues also result in misperceptions of 3D shape (Watt et al. 2005, Hoffman et al. 2008). Figure 15 illustrates that viewing distance compensation is not 100% veridical; we wondered how additional errors in viewing distance estimation resulting from the vergence-accommodation conflict might affect people's ability to maintain depth constancy.

Stimulus and Procedure 3a. We repeated the methods of experiment 2 using a standard grating with a spatial frequency of 0.3cpd, peak-to-trough depth of 4cm, and viewing distance of 1.3D and four comparison gratings with a spatial frequency of 0.3cpd

and vergence distances of either 0.7D or 1.9D. Two gratings had no vergence-accommodation conflict so their accommodation distances were matched to their vergence distance. The other two gratings were presented at a focal distance of 1.3D resulting in either a -0.6D or +0.6D vergence-accommodation conflict (Figure 17). The -0.6D and +0.6D vergence-accommodation trials as well as the 0.7D vergence-accommodation consistent trials were interleaved with experiment 2. Because we did not present 1.9D gratings in experiment 2, we collected the vergence-accommodation consistent results for this condition separately. Image rendering, psychophysical methods, and data analysis were the same as in experiment 2.

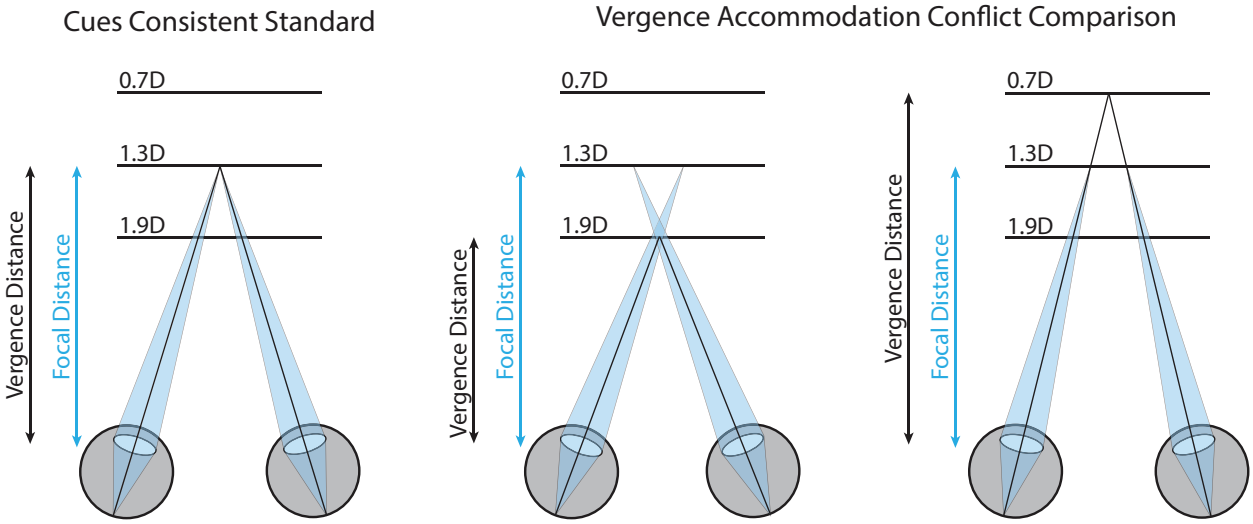


Figure 17 Vergence-accommodation conflict conditions. The left panel shows the vergence-accommodation cues-consistent condition. The central panel shows the -0.6D vergence-accommodation conflict in which the vergence distance is farther than the accommodation distance. The right panel shows the +0.6D vergence-accommodation conflict in which the vergence distance is closer than the accommodation distance.

Stimulus and Procedure 3b. The previous methods in 3a allow us to see if there was a difference between how closely depth constancy is maintained for vergence-accommodation conflict and non-conflict stimuli. We also wanted to directly compare the perceived depth of the conflict and non-conflict waveforms. Thus we repeated the procedures from 3a, but used the 0.7 and 1.9D vergence-accommodation consistent waveforms as the standard with the vergence-accommodation conflict gratings as the comparisons (Figure 18).

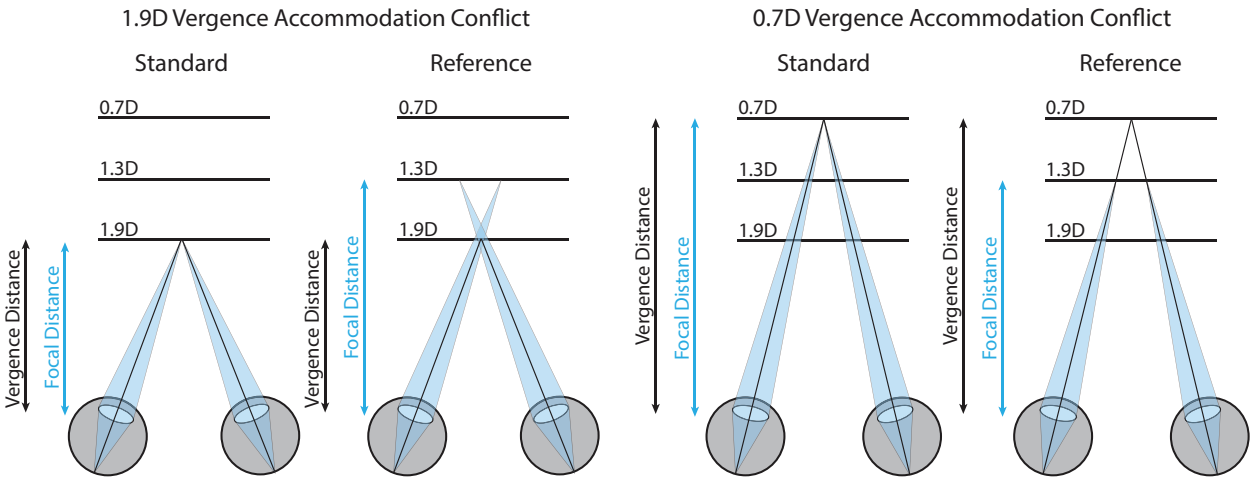


Figure 18 Vergence-accommodation direct comparison conflict conditions. In this experiment we compared the -0.6D (left) vergence-accommodation conflict with a no conflict standard at 1.9D and a +0.6D (right) vergence-accommodation conflict with a no conflict standard at 0.7D.

Subjects. Observers remained unchanged for experiment 3a. We recruited two additional observers for experiment 3b (one male, one female ages 27 and 22). Both had normal stereovision, one had an inter-pupillary distance of 55mm and the other had an inter-pupillary distance of 61mm. Apparatus and data analysis remained unchanged from experiment 2.

Experiment 3: Results

When there is a vergence-accommodation conflict we expect that the estimated viewing distance will be affected by the sign of the conflict. For the +0.6D conflict the estimated viewing distance should be too distant relative to ground truth and for the -0.6D conflict, the viewing distance estimate should be too near. These errors in viewing distance estimation should manifest themselves as errors in the viewing distance disparity compensation. Thus we expect that the +0.6D conflict should result in a disparity that is smaller than the vergence-accommodation consistent comparison and the -0.6D conflict should result in a disparity that is larger than the vergence-accommodation comparison.

Figure 19 shows the results for individual observers and the group average. The dashed lines indicate the disparities of a 4cm peak-to-trough grating at 0.7D, 1.3D, and 1.9D. If depth constancy is maintained then the disparity for the comparisons shown at 1.9D should be at the dashed line labeled 1.9D and the disparity for the comparisons shown at 0.7D should be at the dashed line labeled 0.7D. For the two comparisons with vergence distances of 1.9D we found no difference between the +0.6D conflict and no-conflict condition. For -0.6D conflict there was a difference in perceived depth in the expected direction. It's possible that the weighting of vertical disparity, vergence, and accommodation changes based on viewing distance. In the -0.6D conflict condition the viewing distance is farther away and the vertical disparity signal and horizontal disparity signals are smaller in magnitude. It is possible that under these conditions the weighting given to accommodation to estimate viewing distance estimation increases so the vergence-accommodation conflict has a larger effect on the perceived depth. It would be

interesting to see how these results might change if we had examined conflicts larger than 0.6D.

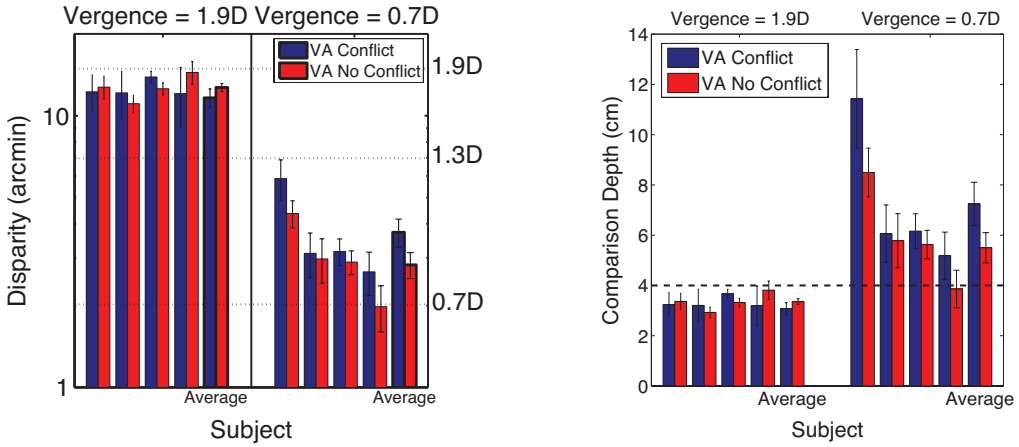


Figure 19 Perceived depth with and without vergence-accommodation conflict. **Left:** Data is plotted in disparity. **Right:** Data is plotted in metric units. The predicted effect of the vergence-accommodation conflict is to the perceived viewing distance from 1.9D to 1.3D and reduces the perceived viewing distance from 0.7D to 1.3D. Thus the cues consistent perceived depth at the 1.9D vergence distance should be larger than the conflict condition and less than the conflict at 0.7D. When the vergence distance is 1.9D the vergence-accommodation conflict appears to have a small, but statistically insignificant effect. At 0.7D vergence distance the conflict has a much larger effect in the predicted direction.

While the group average shows a significant difference between the conflict and no-conflict conditions for the 0.7D vergence condition, two of the four subjects showed no significant difference between the two conditions. We decided to directly compare the perceived depth of a vergence-accommodation conflict stimulus to non-conflict vergence-accommodation stimulus to see if they are perceived as having different depths. Figure 20 shows the results; when the vergence-accommodation consistent standard is at 1.9D, the cue conflict stimulus is perceived to have more depth than intended so a grating with less depth than 4cm appears to match a cues consistent standard with 4cm depth. When the vergence-accommodation consistent standard is at 0.7D the cue conflict stimulus is perceived to have less depth than it should and a grating more than 4cm of depth is perceived to have the same depth as a 4cm standard at 0.7D. These findings show that there is a small, but perceptible effect of the vergence-accommodation conflict on perceived depth with a vergence-accommodation conflict as small as 0.6D. When the intended presentation distance is in front of the display distance depth is over perceived and when the intended presentation is in front of the screen depth is under perceived.

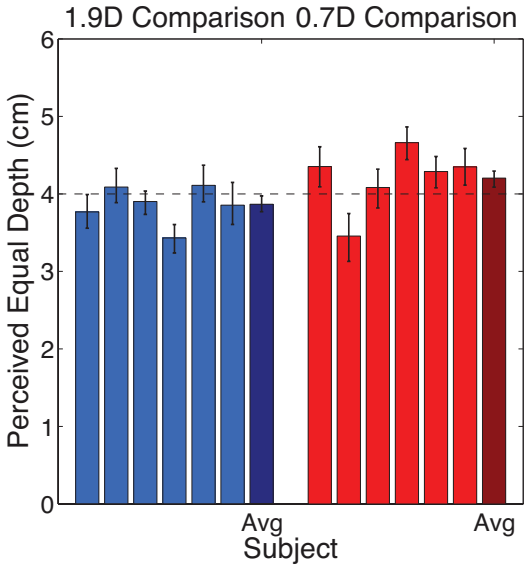


Figure 20 Illustration of the effect of the vergence-accommodation conflict on perceived depth. Individual subject data is shown as the light blue and light red bars while dark blue and dark red represent the averaged group data. Blue bars show the perceived depth of the vergence-accommodation conflict (vergence = 1.9D, accommodation = 1.3D) when the standard was at 1.9D, red bars show the perceived depth of the vergence-accommodation conflict (vergence = 0.7D, accommodation = 1.3D) when the standard was at 0.7D. Overall the vergence-accommodation conflict affects the perceived depth of the conflict stimulus; depth is over perceived when the conflict is in front of the display and under perceived when the conflict is behind the display.

Discussion

Use of JNDs as Perceptible Units in Disparity Compression

With the rise of various displays capable of displaying stereo3D imagery it is helpful to consider to display methods that minimize visual discomfort (Shibata et al. 2011) and maximize visual fidelity (Held and Banks 2008). There is also interest in developing stereo3D compression algorithms to accompany new stereo3D content (Didyk et al. 2011, Pająk et al. 2014). Like previous video algorithms new stereo compression methods aim to incorporate aspects of human vision to preserve perceptible features while discarding imperceptible information. Our results provide insights towards how disparities at different spatial frequencies should be encoded. Didyk et al. (2011) proposed using perceptible units (JNDs) to encode disparities in a perceptually linear space similar to encoding luminance using a gamma function. Disparities could be manipulated based on those JNDs compression, for example, could be easily achieved by eliminating all JDNs < 1 without impacting the perceived depth in the images. Another application of such a procedure is the ability to rescale the depth to restrict vergence to a certain volume relative to the display focal plane. The transducer functions were obtained by measuring the JNDs for sinusoidal disparity gratings at different spatial frequencies and disparity magnitudes and four examples are show in Figure 21.

Our results in Figure 8 and Figure 16 agree with the shape of the transducer functions. There appears to be a spatial frequency dependency on the JNDs and the absolute value of the JNDs increases in disparity for larger disparity magnitudes (our

results are reported in log units). Figure 5 shows the perceived depth from disparity is constant across spatial frequencies above ~4arcmin, this means any compression or expansion of disparity in perceptually linear JND space should result in the same disparity magnitude across all spatial frequencies after conversion back into disparity units. Figure 21B shows that this is indeed the case, at 50arcmin the JNDs for the four spatial frequency bands are shown with solid lines. We reduce their JNDs by ½ and find that the resulting disparities are very similar as desired. To use the above method in compression Didyak et al. simply discarded JNDs below a certain value. For JNDs < 1 this is corresponds to eliminating sub-thresholds disparities which should have no effect on the depth percept. For JNDs >1, however, this method will result in depth artifacts because this affects the disparity magnitudes at each spatial frequency band differently.

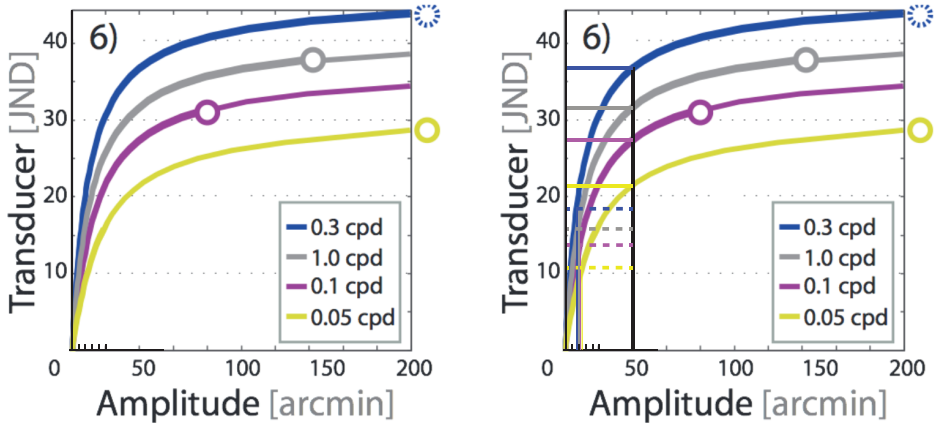


Figure 21 Left: Transducer functions from Didyk et al. (2011) used to linearize perceived depth from disparity. The open circles represent the upper-disparity gradient limit. **Right:** Reducing the JNDs by 50% results in approximately the same disparity for all spatial frequency bands. Solid bars indicate JND units at 50arcmin for all four spatial frequencies; the dashed lines represent a 50% reduction in JND units. The new disparity amplitudes for at each of the four spatial frequencies after disparity scaling are all approximately 10arcmin +/- 2arcmin.

Vergence-Accommodation Conflict in Stereo3D Displays

Our data from experiment 2 show that perceived depth from disparity is highly dependent on viewing distance. Experiment 3 shows that the vergence-accommodation conflict has the potential to interfere with veridical viewing distance estimation and can lead to misperceptions in depth from disparity. Many of the head-mounted displays (HMDs) designed for virtual and augmented reality display stereo3D content at a fixed focal plane resulting in a vergence-accommodation conflict. The decision for the placement of the display should be made to maximize comfort and depth fidelity. The zone of comfort (Shibatta et al. 2011) depends on the viewing distance, a focal distance of 2D allows for a comfortable viewing volume from 0.2D to 3.2D or ~31cm to 5m. From experiment 3 it is appears that conflicts as small as 0.6D can result misperceptions in depth. However this effect may also depend on the direction of the conflict as well as the disparity magnitude. If larger vergence-accommodation conflicts do result in depth misperceptions, there may be an optimal viewing distance for comfort and depth fidelity that is more restrictive than the zone of comfort.

Misperception of Depth from Errors in Inter-Pupillary Distance

We've shown that proper viewing distance estimation is critical to perceive the correct depth from disparity, but what happens with the inter-pupillary distance of the observer does not match the intended inter-pupillary distance used to render the stereo content? Traditional stereo3D display like a TV or movie theater screen displays the same set of on-screen disparities to all observers. A personal HMD requires the IPD to be properly calibrated for the user. In both cases there is the potential to show subjects disparities that were rendered for an IPD that does not match the user's actual IPD. We used the JNDs from experiment 1 to determine how large of an error could be tolerated for an inter-pupillary distance of 56mm, 63mm, and 71mm which corresponds to the 5th, 50th, and 95th percentile (Dodgson 2004). The blue bars illustrate the JNDs obtained for three disparity magnitudes (4, 16, and 32arcmin) from experiment 1. The JNDs nearly reach the disparities obtained calculating disparity for the bottom 5th and top 95th percentile inter-pupillary distances. This suggests that large errors in inter-pupillary distance can be tolerated without greatly affecting perceived depth.

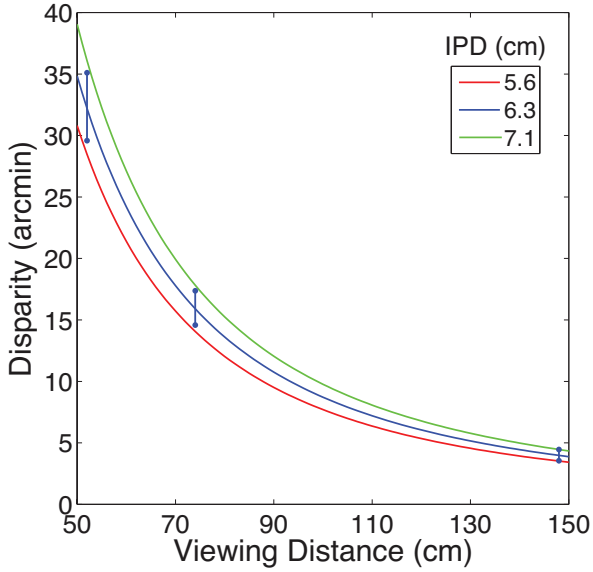


Figure 22 Disparities calculated for an object 4cm away from fixation from viewing distances of 50 to 150cm. The red, blue, and green lines represent the disparity calculated using an inter-pupillary distance corresponding to the 5th, 50th, and 95th percentile inter-pupillary distances. The blue dots represent the JND for disparity at 4, 16, and 32arcmin from experiment 1. Any points within the blue region appear to have the same depth thus using a 63mm inter-pupillary assumption is might be sufficiently accurate for a vast majority of the population.

Conclusion

We found that depth constancy occurs across a broad range of corrugation frequencies provided that the disparity amplitude is not close to the lower-disparity threshold or upper-disparity limit. Low spatial frequencies avoid the disparity-gradient limit at all but the greatest amplitudes and can convey more apparent depth than other frequencies. It is necessary to scale depth from disparity based on fixation distance, which can be estimated using a combination of vergence, accommodation, and vertical disparity.

The visual system is able to make the appropriate depth scaling compensation at different viewing distances and spatial frequencies. When there is a vergence-accommodation conflict viewing distance compensation is affected and depth can be misperceived. These results are the first to demonstrate depth constancy across spatial frequencies as well as across spatial frequency and viewing distance simultaneously.

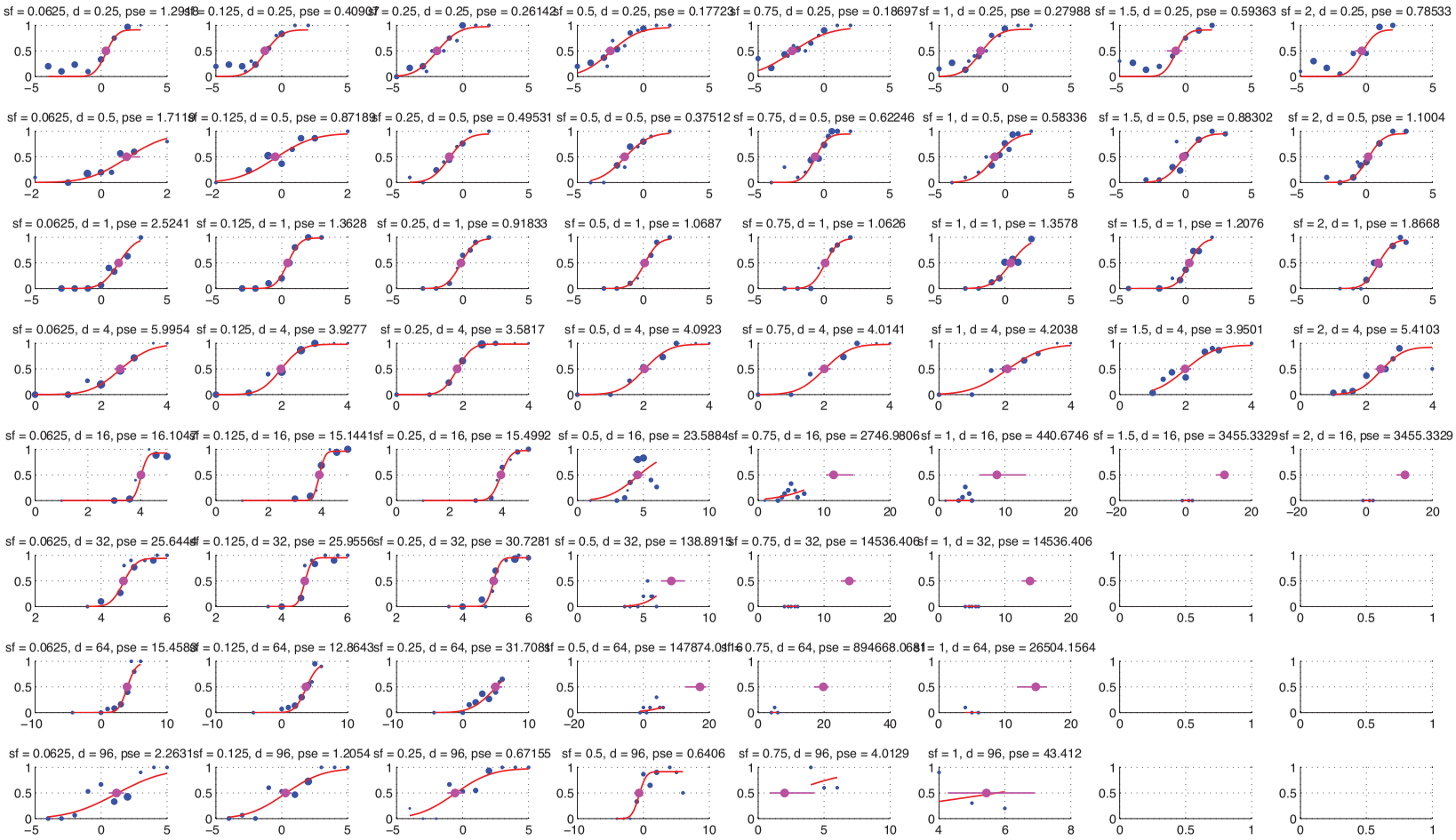
Appendix

Experiment 1: Psychometric Functions

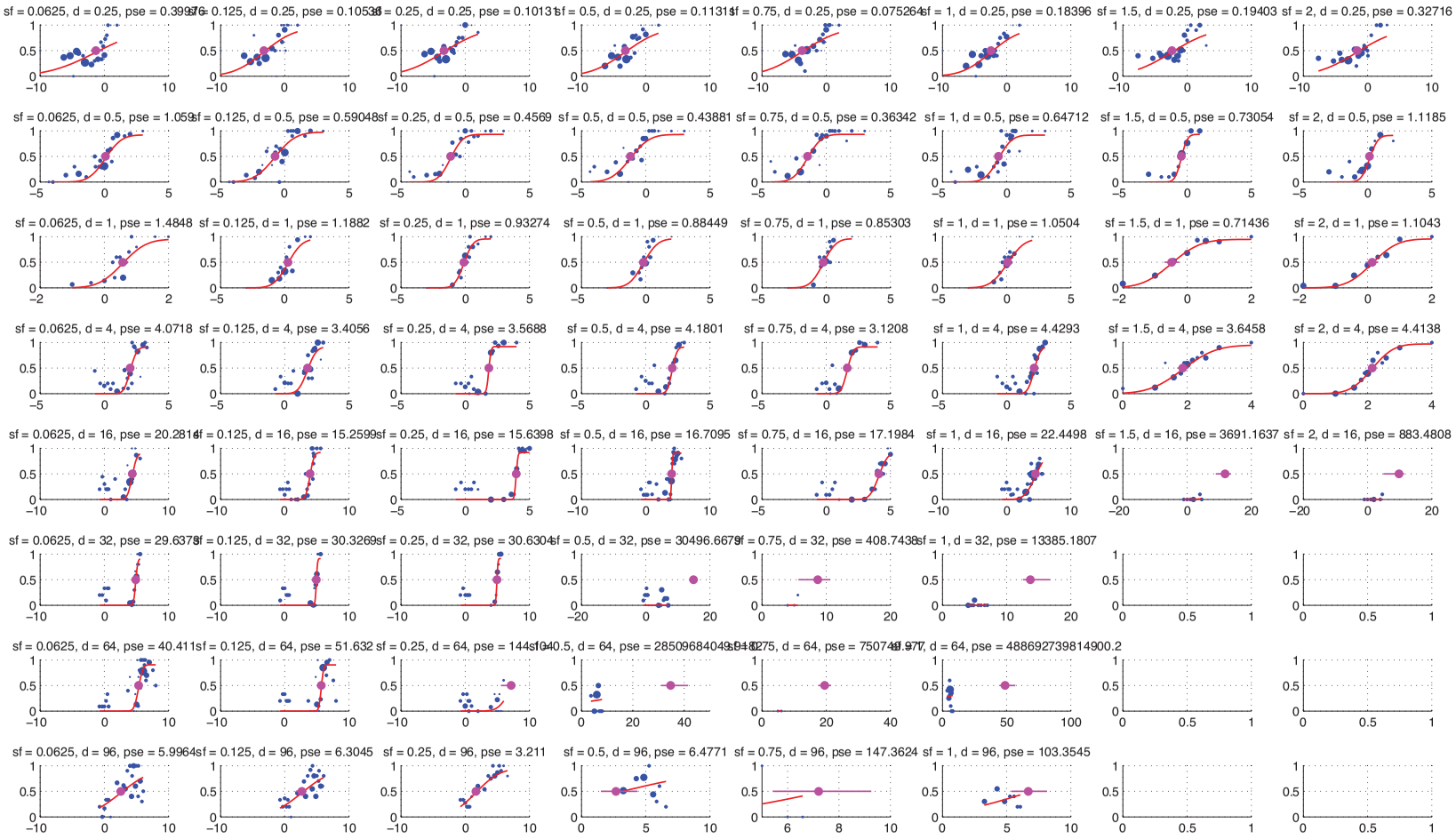
The psychometric functions from experiment 1 for each observer and the combined data across all subjects are shown in the following pages.

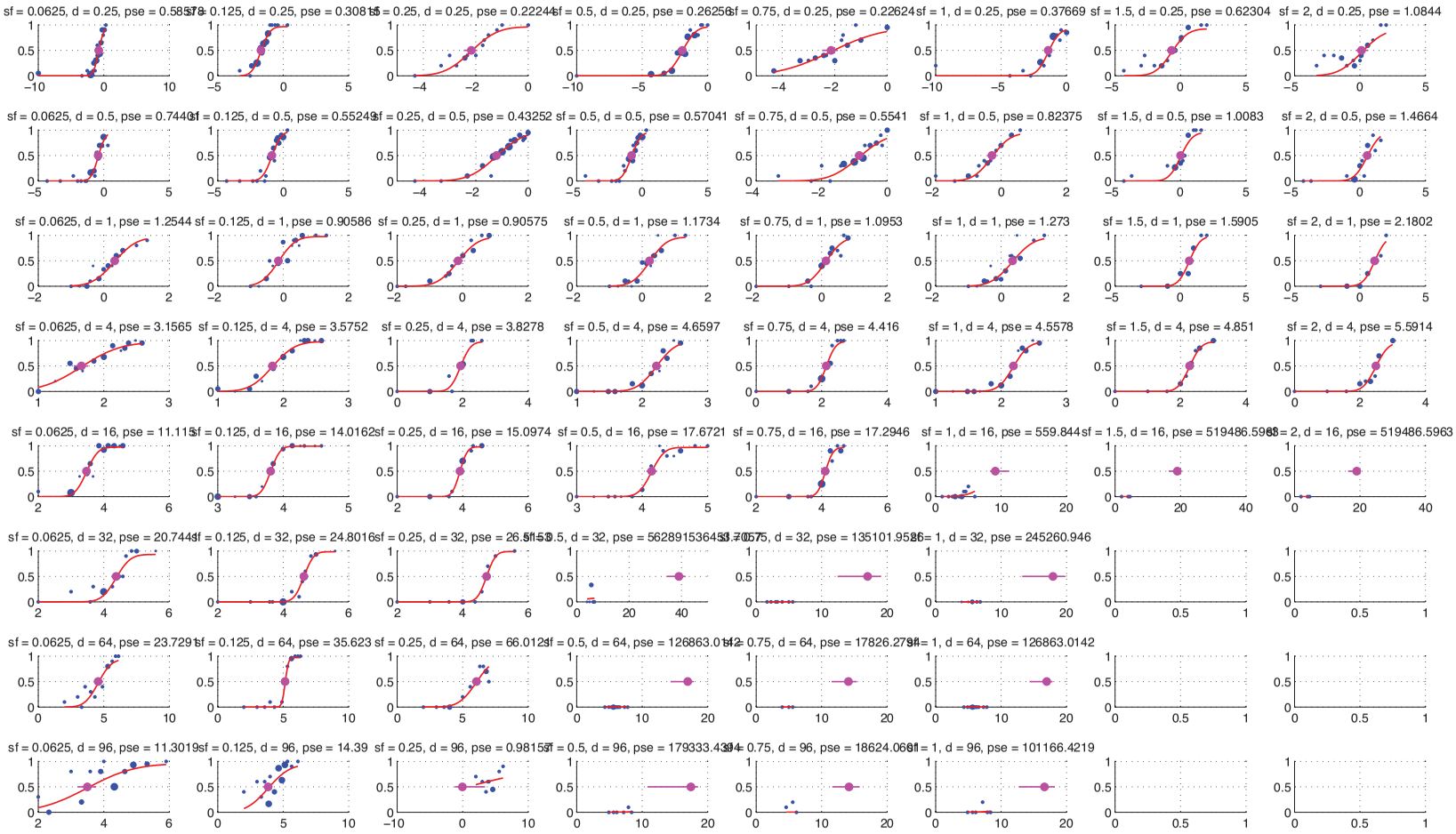
Group Average

Subject 1

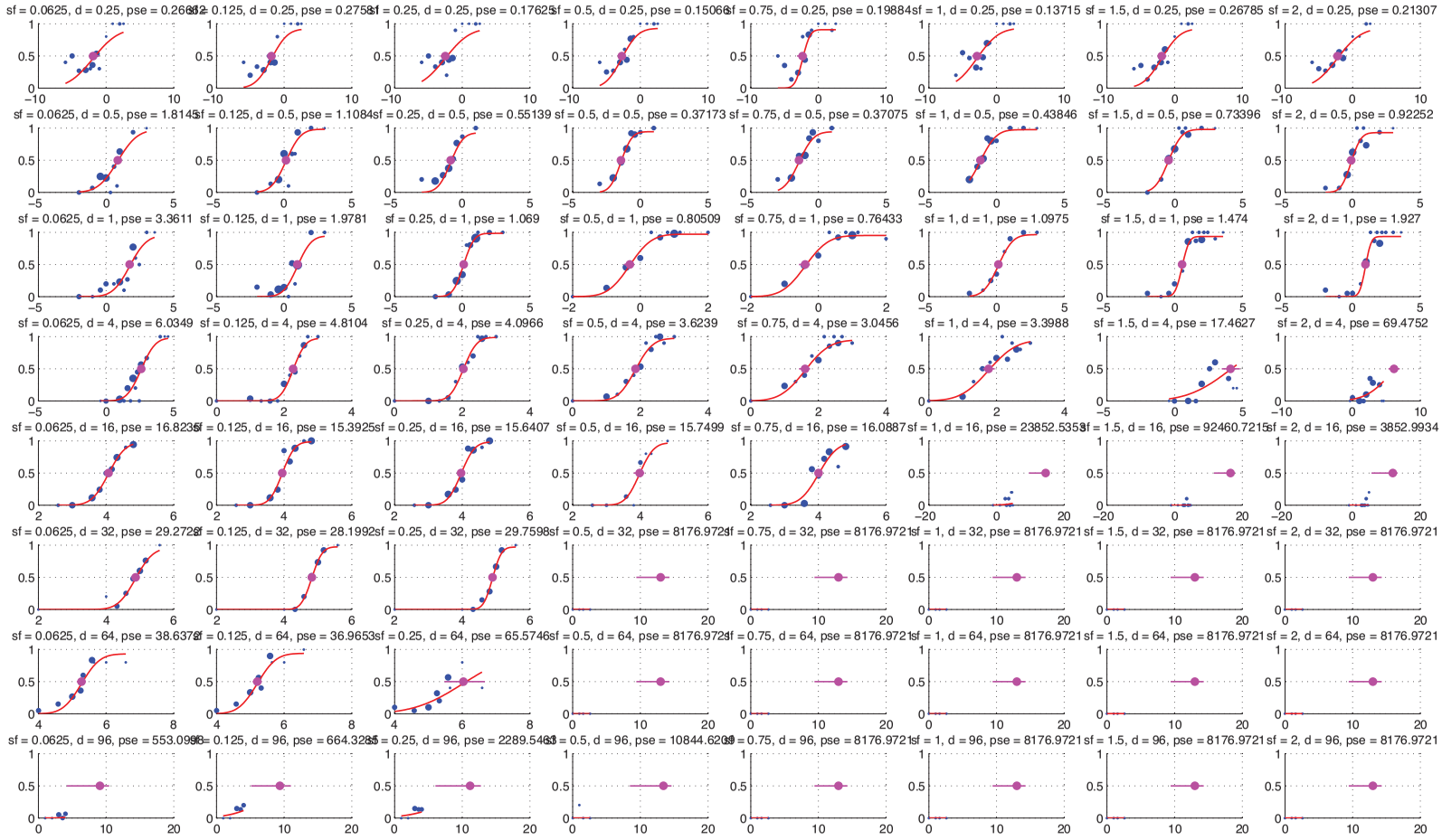

sf = 0.0625, d = 0.25, pse = 1.2998
sf = 0.125, d = 0.25, pse = 0.40967
sf = 0.25, d = 0.25, pse = 0.26142
sf = 0.5, d = 0.25, pse = 0.17723
sf = 0.75, d = 0.25, pse = 0.18697
sf = 1, d = 0.25, pse = 0.27988
sf = 1.5, d = 0.25, pse = 0.59363
sf = 2, d = 0.25, pse = 0.78533
sf = 0.0625, d = 0.5, pse = 1.7115
sf = 0.125, d = 0.5, pse = 0.87189
sf = 0.25, d = 0.5, pse = 0.49531
sf = 0.5, d = 0.5, pse = 0.37512
sf = 0.75, d = 0.5, pse = 0.62246
sf = 1, d = 0.5, pse = 0.58336
sf = 1.5, d = 0.5, pse = 0.88302
sf = 2, d = 0.5, pse = 1.1004
sf = 0.0625, d = 1, pse = 2.5241
sf = 0.125, d = 1, pse = 1.3628
sf = 0.25, d = 1, pse = 0.91833
sf = 0.5, d = 1, pse = 1.0687
sf = 0.75, d = 1, pse = 1.0626
sf = 1, d = 1, pse = 1.3578
sf = 1.5, d = 1, pse = 1.2076
sf = 2, d = 1, pse = 1.8668
sf = 0.0625, d = 4, pse = 5.9954
sf = 0.125, d = 4, pse = 3.9277
sf = 0.25, d = 4, pse = 3.5817
sf = 0.5, d = 4, pse = 4.0923
sf = 0.75, d = 4, pse = 4.0141
sf = 1, d = 4, pse = 4.2038
sf = 1.5, d = 4, pse = 3.9501
sf = 2, d = 4, pse = 5.4103
sf = 0.0625, d = 16, pse = 16.1043
sf = 0.125, d = 16, pse = 15.1441
sf = 0.25, d = 16, pse = 15.4992
sf = 0.5, d = 16, pse = 23.5884
sf = 0.75, d = 16, pse = 2746.9806
sf = 1, d = 16, pse = 440.6746
sf = 1.5, d = 16, pse = 3455.3329
sf = 2, d = 16, pse = 3455.3329
sf = 0.0625, d = 32, pse = 25.6446
sf = 0.125, d = 32, pse = 25.9556
sf = 0.25, d = 32, pse = 30.7281
sf = 0.5, d = 32, pse = 138.8915
sf = 0.75, d = 32, pse = 14536.406
sf = 1, d = 32, pse = 14536.406
sf = 0.0625, d = 64, pse = 15.4588
sf = 0.125, d = 64, pse = 12.8643
sf = 0.25, d = 64, pse = 31.7083
sf = 0.5, d = 64, pse = 147874.0816
sf = 0.75, d = 64, pse = 894668.0681
sf = 1, d = 64, pse = 26504.1564
sf = 0.0625, d = 96, pse = 2.2631
sf = 0.125, d = 96, pse = 1.2054
sf = 0.25, d = 96, pse = 0.67155
sf = 0.5, d = 96, pse = 0.6406
sf = 0.75, d = 96, pse = 4.0129
sf = 1, d = 96, pse = 43.412

Subject 2


sf = 0.0625, d = 0.25, pse = 0.39976
sf = 0.125, d = 0.25, pse = 0.10536
sf = 0.25, d = 0.25, pse = 0.10131
sf = 0.5, d = 0.25, pse = 0.11315
sf = 0.75, d = 0.25, pse = 0.075264
sf = 1, d = 0.25, pse = 0.18396
sf = 1.5, d = 0.25, pse = 0.19403
sf = 2, d = 0.25, pse = 0.32716
sf = 0.0625, d = 0.5, pse = 1.059
sf = 0.125, d = 0.5, pse = 0.59048
sf = 0.25, d = 0.5, pse = 0.4569
sf = 0.5, d = 0.5, pse = 0.43881
sf = 0.75, d = 0.5, pse = 0.36342
sf = 1, d = 0.5, pse = 0.64712
sf = 1.5, d = 0.5, pse = 0.73054
sf = 2, d = 0.5, pse = 1.1185
sf = 0.0625, d = 1, pse = 1.4848
sf = 0.125, d = 1, pse = 1.1882
sf = 0.25, d = 1, pse = 0.93274
sf = 0.5, d = 1, pse = 0.88449
sf = 0.75, d = 1, pse = 0.85303
sf = 1, d = 1, pse = 1.0504
sf = 1.5, d = 1, pse = 0.71436
sf = 2, d = 1, pse = 1.1043
sf = 0.0625, d = 4, pse = 4.0718
sf = 0.125, d = 4, pse = 3.4056
sf = 0.25, d = 4, pse = 3.5688
sf = 0.5, d = 4, pse = 4.1801
sf = 0.75, d = 4, pse = 3.1208
sf = 1, d = 4, pse = 4.4293
sf = 1.5, d = 4, pse = 3.6458
sf = 2, d = 4, pse = 4.4138
sf = 0.0625, d = 16, pse = 20.2814
sf = 0.125, d = 16, pse = 15.2599
sf = 0.25, d = 16, pse = 15.6398
sf = 0.5, d = 16, pse = 16.7095
sf = 0.75, d = 16, pse = 17.1984
sf = 1, d = 16, pse = 22.4498
sf = 1.5, d = 16, pse = 3691.1637
sf = 2, d = 16, pse = 883.4808
sf = 0.0625, d = 32, pse = 29.6378
sf = 0.125, d = 32, pse = 30.3269
sf = 0.25, d = 32, pse = 30.6304
sf = 0.5, d = 32, pse = 30496.6679
sf = 0.75, d = 32, pse = 408.7438
sf = 1, d = 32, pse = 13385.1807
sf = 0.0625, d = 64, pse = 40.411
sf = 0.125, d = 64, pse = 51.632
sf = 0.25, d = 64, pse = 144.9
sf = 0.5, d = 64, pse = 2850968404
sf = 0.75, d = 64, pse = 7507
sf = 1, d = 64, pse = 4886927398
sf = 0.0625, d = 96, pse = 5.9964
sf = 0.125, d = 96, pse = 6.3045
sf = 0.25, d = 96, pse = 3.211
sf = 0.5, d = 96, pse = 6.4771
sf = 0.75, d = 96, pse = 147.3624
sf = 1, d = 96, pse = 103.3545


Subject 3
sf = 0.0625, d = 0.25, pse = 0.58578
sf = 0.125, d = 0.25, pse = 0.30815
sf = 0.25, d = 0.25, pse = 0.22244
sf = 0.5, d = 0.25, pse = 0.26256
sf = 0.75, d = 0.25, pse = 0.22624
sf = 1, d = 0.25, pse = 0.37669
sf = 1.5, d = 0.25, pse = 0.62304
sf = 2, d = 0.25, pse = 1.0844
sf = 0.0625, d = 0.5, pse = 0.74401
sf = 0.125, d = 0.5, pse = 0.55249
sf = 0.25, d = 0.5, pse = 0.43252
sf = 0.5, d = 0.5, pse = 0.57041
sf = 0.75, d = 0.5, pse = 0.5541
sf = 1, d = 0.5, pse = 0.82375
sf = 1.5, d = 0.5, pse = 1.0083
sf = 2, d = 0.5, pse = 1.4664
sf = 0.0625, d = 1, pse = 1.2544
sf = 0.125, d = 1, pse = 0.90586
sf = 0.25, d = 1, pse = 0.90575
sf = 0.5, d = 1, pse = 1.1734
sf = 0.75, d = 1, pse = 1.0953
sf = 1, d = 1, pse = 1.273
sf = 1.5, d = 1, pse = 1.5905
sf = 2, d = 1, pse = 2.1802
sf = 0.0625, d = 4, pse = 3.1565
sf = 0.125, d = 4, pse = 3.5752
sf = 0.25, d = 4, pse = 3.8278
sf = 0.5, d = 4, pse = 4.6597
sf = 0.75, d = 4, pse = 4.416
sf = 1, d = 4, pse = 4.5578
sf = 1.5, d = 4, pse = 4.851
sf = 2, d = 4, pse = 5.5914
sf = 0.0625, d = 16, pse = 11.115
sf = 0.125, d = 16, pse = 14.0162
sf = 0.25, d = 16, pse = 15.0974
sf = 0.5, d = 16, pse = 17.6721
sf = 0.75, d = 16, pse = 17.2946
sf = 1, d = 16, pse = 559.844
sf = 1.5, d = 16, pse = 519486.5963
sf = 2, d = 16, pse = 519486.5963
sf = 0.0625, d = 32, pse = 20.7445
sf = 0.125, d = 32, pse = 24.8016
sf = 0.25, d = 32, pse = 26.5153
sf = 0.5, d = 32, pse = 562891536453
sf = 0.75, d = 32, pse = 135101.9526
sf = 1, d = 32, pse = 245260.946
sf = 0.0625, d = 64, pse = 23.7291
sf = 0.125, d = 64, pse = 35.623
sf = 0.25, d = 64, pse = 66.0123
sf = 0.5, d = 64, pse = 126863.0142
sf = 0.75, d = 64, pse = 17826.2794
sf = 1, d = 64, pse = 126863.0142
sf = 0.0625, d = 96, pse = 11.3019
sf = 0.125, d = 96, pse = 14.39
sf = 0.25, d = 96, pse = 0.98157
sf = 0.5, d = 96, pse = 179333.4394
sf = 0.75, d = 96, pse = 18624.0661
sf = 1, d = 96, pse = 101166.4219

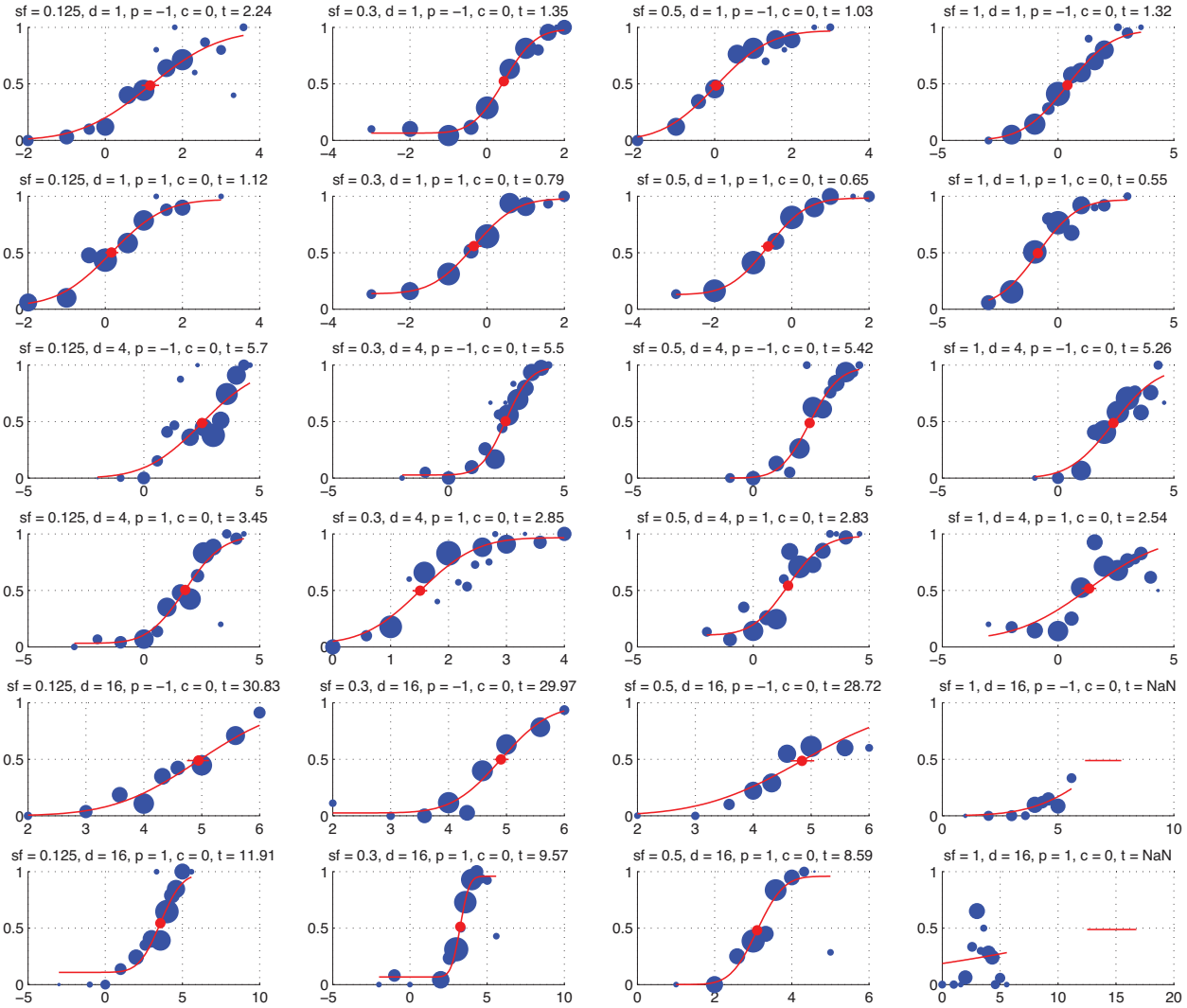
Subject 4


sf = 0.0625, d = 0.25, pse = 0.26682
sf = 0.125, d = 0.25, pse = 0.27581
sf = 0.25, d = 0.25, pse = 0.17625
sf = 0.5, d = 0.25, pse = 0.15066
sf = 0.75, d = 0.25, pse = 0.19884
sf = 1, d = 0.25, pse = 0.13715
sf = 1.5, d = 0.25, pse = 0.26785
sf = 2, d = 0.25, pse = 0.21307
sf = 0.0625, d = 0.5, pse = 1.8145
sf = 0.125, d = 0.5, pse = 1.1084
sf = 0.25, d = 0.5, pse = 0.55139
sf = 0.5, d = 0.5, pse = 0.37173
sf = 0.75, d = 0.5, pse = 0.37075
sf = 1, d = 0.5, pse = 0.43846
sf = 1.5, d = 0.5, pse = 0.73396
sf = 2, d = 0.5, pse = 0.92252
sf = 0.0625, d = 1, pse = 3.3611
sf = 0.125, d = 1, pse = 1.9781
sf = 0.25, d = 1, pse = 1.069
sf = 0.5, d = 1, pse = 0.80509
sf = 0.75, d = 1, pse = 0.76433
sf = 1, d = 1, pse = 1.0975
sf = 1.5, d = 1, pse = 1.474
sf = 2, d = 1, pse = 1.927
sf = 0.0625, d = 4, pse = 6.0349
sf = 0.125, d = 4, pse = 4.8104
sf = 0.25, d = 4, pse = 4.0966
sf = 0.5, d = 4, pse = 3.6239
sf = 0.75, d = 4, pse = 3.0456
sf = 1, d = 4, pse = 3.3988
sf = 1.5, d = 4, pse = 17.4627
sf = 2, d = 4, pse = 69.4752
sf = 0.0625, d = 16, pse = 16.8235
sf = 0.125, d = 16, pse = 15.3925
sf = 0.25, d = 16, pse = 15.6407
sf = 0.5, d = 16, pse = 15.7499
sf = 0.75, d = 16, pse = 16.0887
sf = 1, d = 16, pse = 23852.5353
sf = 1.5, d = 16, pse = 92460.7215
sf = 2, d = 16, pse = 3852.9934
sf = 0.0625, d = 32, pse = 29.2726
sf = 0.125, d = 32, pse = 28.1992
sf = 0.25, d = 32, pse = 29.7598
sf = 0.5, d = 32, pse = 8176.9721
sf = 0.75, d = 32, pse = 8176.9721
sf = 1, d = 32, pse = 8176.9721
sf = 1.5, d = 32, pse = 8176.9721
sf = 2, d = 32, pse = 8176.9721
sf = 0.0625, d = 64, pse = 38.6378
sf = 0.125, d = 64, pse = 36.9653
sf = 0.25, d = 64, pse = 65.5746
sf = 0.5, d = 64, pse = 8176.9721
sf = 0.75, d = 64, pse = 8176.9721
sf = 1, d = 64, pse = 8176.9721
sf = 1.5, d = 64, pse = 8176.9721
sf = 2, d = 64, pse = 8176.9721
sf = 0.0625, d = 96, pse = 553.0998
sf = 0.125, d = 96, pse = 664.3285
sf = 0.25, d = 96, pse = 2289.5463
sf = 0.5, d = 96, pse = 10844.6209
sf = 0.75, d = 96, pse = 8176.9721
sf = 1, d = 96, pse = 8176.9721
sf = 1.5, d = 96, pse = 8176.9721
sf = 2, d = 96, pse = 8176.9721

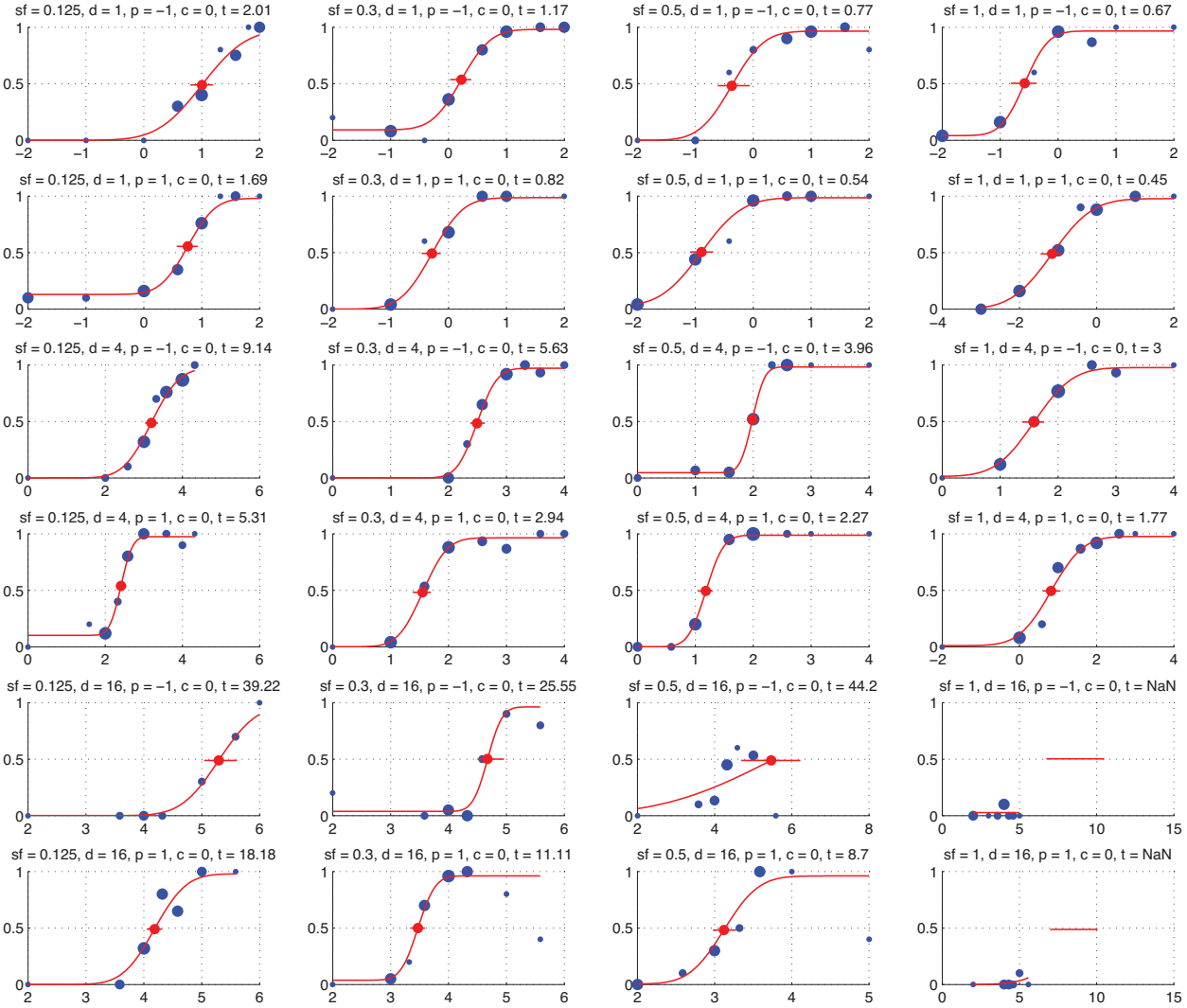
Experiment 2: Psychometric Functions

The psychometric functions from experiment 2 for each observer and the combined data across all subjects are shown in the following pages.

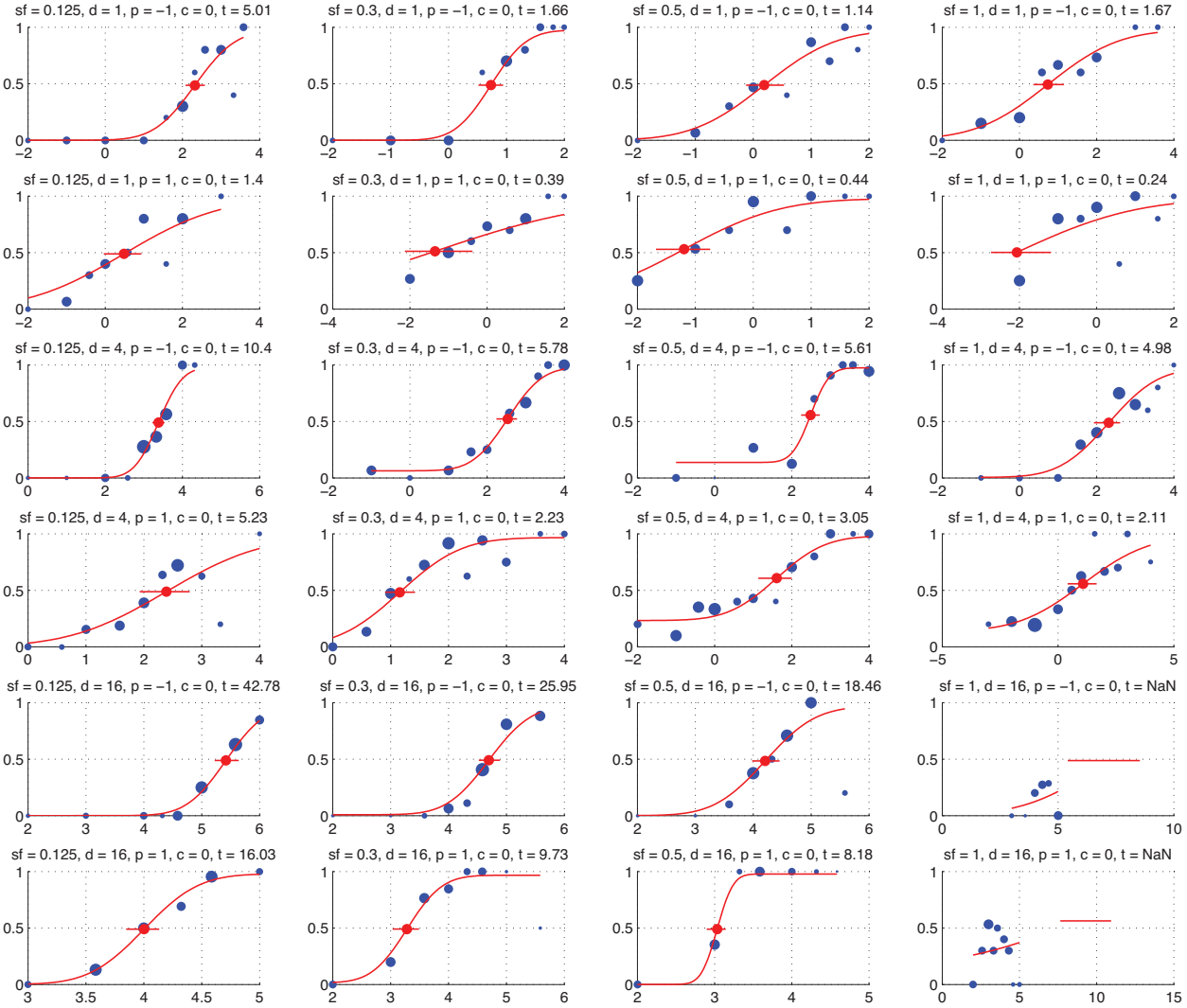
Group Average



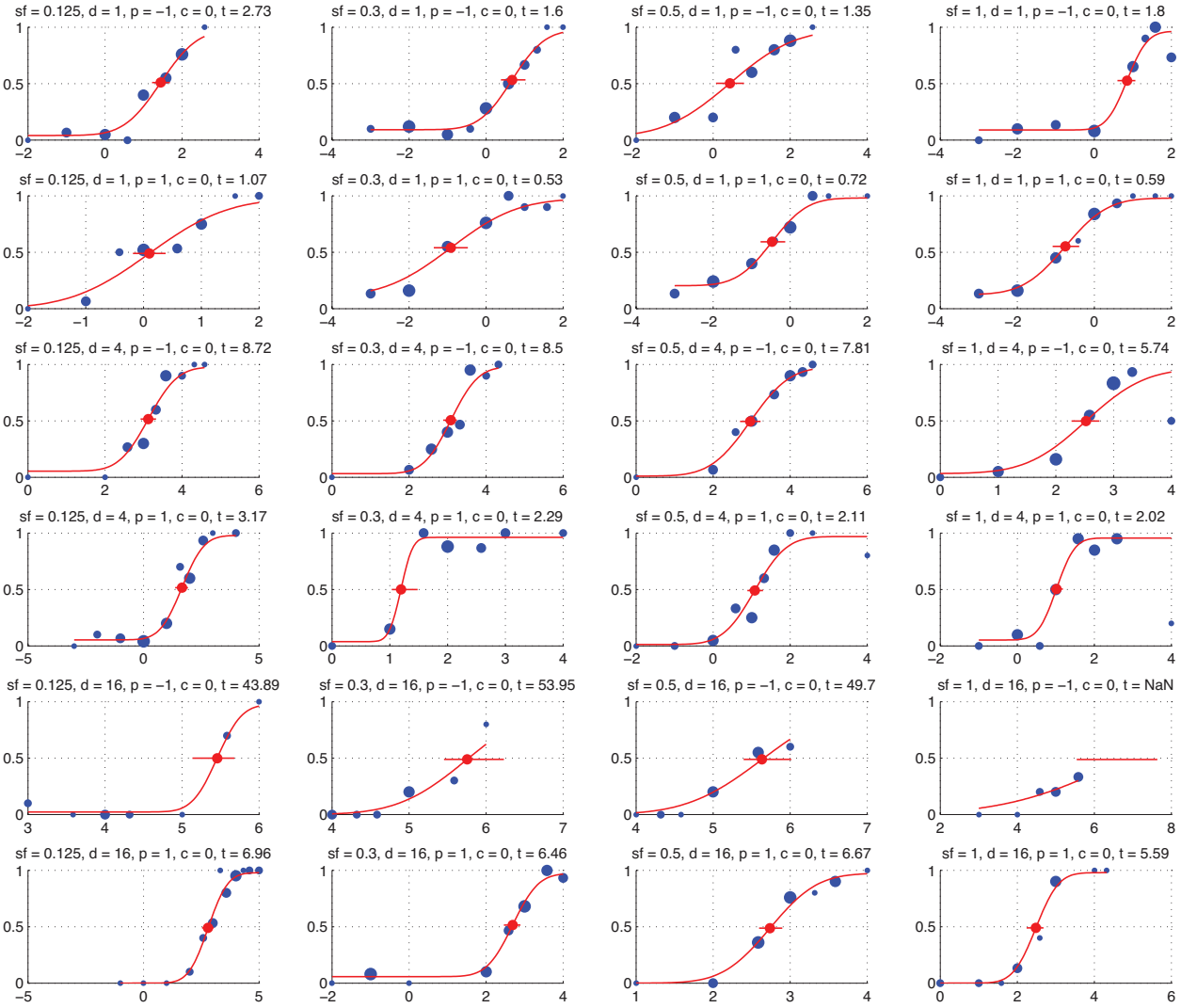
Subject 1


sf = 0.125, d = 1, p = −1, c = 0, t = 2.01
sf = 0.3, d = 1, p = −1, c = 0, t = 1.17
sf = 0.5, d = 1, p = −1, c = 0, t = 0.77
sf = 1, d = 1, p = −1, c = 0, t = 0.67
sf = 0.125, d = 1, p = 1, c = 0, t = 1.69
sf = 0.3, d = 1, p = 1, c = 0, t = 0.82
sf = 0.5, d = 1, p = 1, c = 0, t = 0.54
sf = 1, d = 1, p = 1, c = 0, t = 0.45
sf = 0.125, d = 4, p = −1, c = 0, t = 9.14
sf = 0.3, d = 4, p = −1, c = 0, t = 5.63
sf = 0.5, d = 4, p = −1, c = 0, t = 3.96
sf = 1, d = 4, p = −1, c = 0, t = 3
sf = 0.125, d = 4, p = 1, c = 0, t = 5.31
sf = 0.3, d = 4, p = 1, c = 0, t = 2.94
sf = 0.5, d = 4, p = 1, c = 0, t = 2.27
sf = 1, d = 4, p = 1, c = 0, t = 1.77
sf = 0.125, d = 16, p = −1, c = 0, t = 39.22
sf = 0.3, d = 16, p = −1, c = 0, t = 25.55
sf = 0.5, d = 16, p = −1, c = 0, t = 44.2
sf = 1, d = 16, p = −1, c = 0, t = NaN
sf = 0.125, d = 16, p = 1, c = 0, t = 18.18
sf = 0.3, d = 16, p = 1, c = 0, t = 11.11
sf = 0.5, d = 16, p = 1, c = 0, t = 8.7
sf = 1, d = 16, p = 1, c = 0, t = NaN

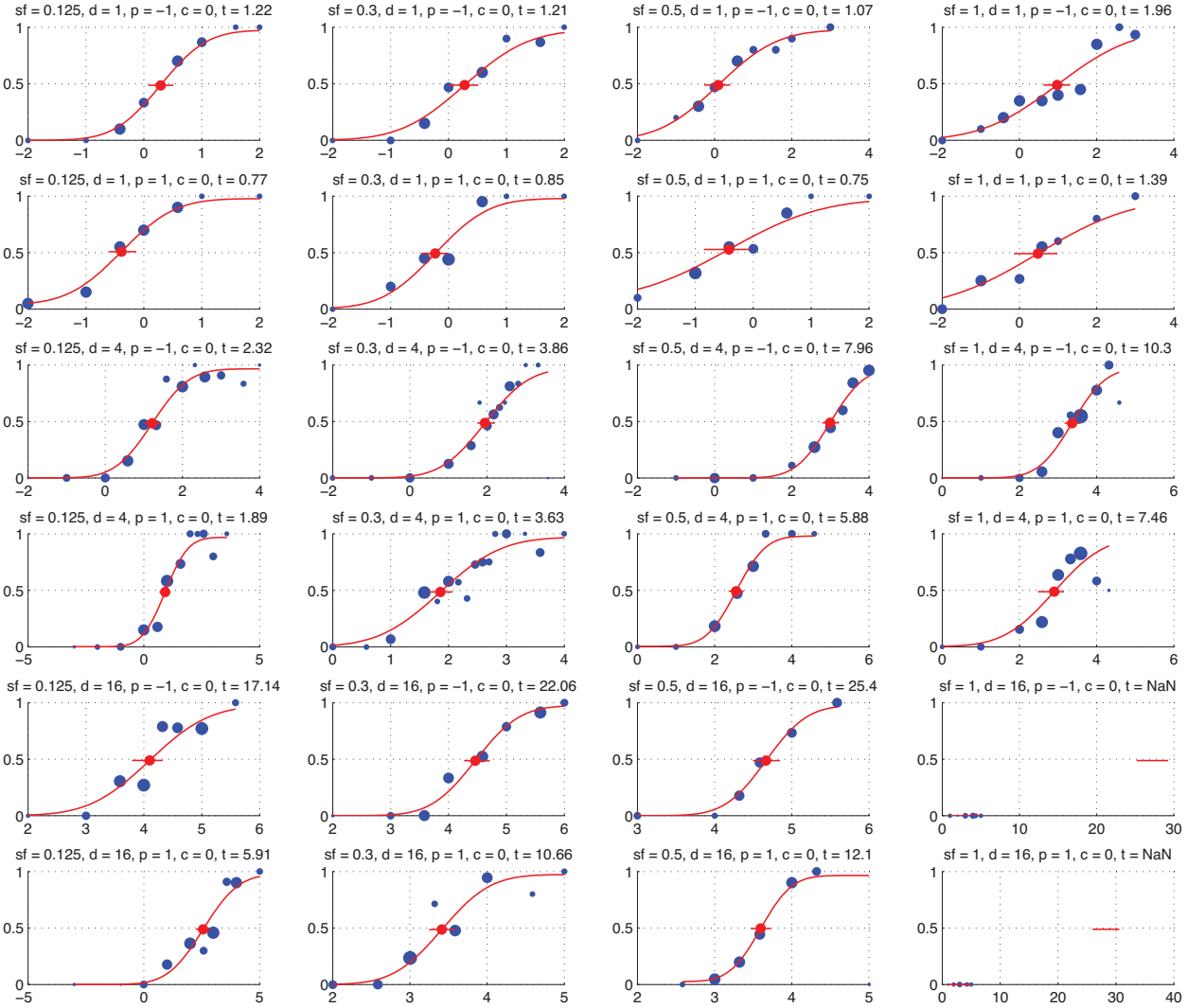
Subject 2


sf = 0.125, d = 1, p = −1, c = 0, t = 5.01
sf = 0.3, d = 1, p = −1, c = 0, t = 1.66
sf = 0.5, d = 1, p = −1, c = 0, t = 1.14
sf = 1, d = 1, p = −1, c = 0, t = 1.67
sf = 0.125, d = 1, p = 1, c = 0, t = 1.4
sf = 0.3, d = 1, p = 1, c = 0, t = 0.39
sf = 0.5, d = 1, p = 1, c = 0, t = 0.44
sf = 1, d = 1, p = 1, c = 0, t = 0.24
sf = 0.125, d = 4, p = −1, c = 0, t = 10.4
sf = 0.3, d = 4, p = −1, c = 0, t = 5.78
sf = 0.5, d = 4, p = −1, c = 0, t = 5.61
sf = 1, d = 4, p = −1, c = 0, t = 4.98
sf = 0.125, d = 4, p = 1, c = 0, t = 5.23
sf = 0.3, d = 4, p = 1, c = 0, t = 2.23
sf = 0.5, d = 4, p = 1, c = 0, t = 3.05
sf = 1, d = 4, p = 1, c = 0, t = 2.11
sf = 0.125, d = 16, p = −1, c = 0, t = 42.78
sf = 0.3, d = 16, p = −1, c = 0, t = 25.95
sf = 0.5, d = 16, p = −1, c = 0, t = 18.46
sf = 1, d = 16, p = −1, c = 0, t = NaN
sf = 0.125, d = 16, p = 1, c = 0, t = 16.03
sf = 0.3, d = 16, p = 1, c = 0, t = 9.73
sf = 0.5, d = 16, p = 1, c = 0, t = 8.18
sf = 1, d = 16, p = 1, c = 0, t = NaN

Subject 3


sf = 0.125, d = 1, p = −1, c = 0, t = 2.73
sf = 0.3, d = 1, p = −1, c = 0, t = 1.6
sf = 0.5, d = 1, p = −1, c = 0, t = 1.35
sf = 1, d = 1, p = −1, c = 0, t = 1.8
sf = 0.125, d = 1, p = 1, c = 0, t = 1.07
sf = 0.3, d = 1, p = 1, c = 0, t = 0.53
sf = 0.5, d = 1, p = 1, c = 0, t = 0.72
sf = 1, d = 1, p = 1, c = 0, t = 0.59
sf = 0.125, d = 4, p = −1, c = 0, t = 8.72
sf = 0.3, d = 4, p = −1, c = 0, t = 8.5
sf = 0.5, d = 4, p = −1, c = 0, t = 7.81
sf = 1, d = 4, p = −1, c = 0, t = 5.74
sf = 0.125, d = 4, p = 1, c = 0, t = 3.17
sf = 0.3, d = 4, p = 1, c = 0, t = 2.29
sf = 0.5, d = 4, p = 1, c = 0, t = 2.11
sf = 1, d = 4, p = 1, c = 0, t = 2.02
sf = 0.125, d = 16, p = −1, c = 0, t = 43.89
sf = 0.3, d = 16, p = −1, c = 0, t = 53.95
sf = 0.5, d = 16, p = −1, c = 0, t = 49.7
sf = 1, d = 16, p = −1, c = 0, t = NaN
sf = 0.125, d = 16, p = 1, c = 0, t = 6.96
sf = 0.3, d = 16, p = 1, c = 0, t = 6.46
sf = 0.5, d = 16, p = 1, c = 0, t = 6.67
sf = 1, d = 16, p = 1, c = 0, t = 5.59

Subject 4


sf = 0.125, d = 1, p = −1, c = 0, t = 1.22
sf = 0.3, d = 1, p = −1, c = 0, t = 1.21
sf = 0.5, d = 1, p = −1, c = 0, t = 1.07
sf = 1, d = 1, p = −1, c = 0, t = 1.96
sf = 0.125, d = 1, p = 1, c = 0, t = 0.77
sf = 0.3, d = 1, p = 1, c = 0, t = 0.85
sf = 0.5, d = 1, p = 1, c = 0, t = 0.75
sf = 1, d = 1, p = 1, c = 0, t = 1.39
sf = 0.125, d = 4, p = −1, c = 0, t = 2.32
sf = 0.3, d = 4, p = −1, c = 0, t = 3.86
sf = 0.5, d = 4, p = −1, c = 0, t = 7.96
sf = 1, d = 4, p = −1, c = 0, t = 10.3
sf = 0.125, d = 4, p = 1, c = 0, t = 1.89
sf = 0.3, d = 4, p = 1, c = 0, t = 3.63
sf = 0.5, d = 4, p = 1, c = 0, t = 5.88
sf = 1, d = 4, p = 1, c = 0, t = 7.46
sf = 0.125, d = 16, p = −1, c = 0, t = 17.14
sf = 0.3, d = 16, p = −1, c = 0, t = 22.06
sf = 0.5, d = 16, p = −1, c = 0, t = 25.4
sf = 1, d = 16, p = −1, c = 0, t = NaN
sf = 0.125, d = 16, p = 1, c = 0, t = 5.91
sf = 0.3, d = 16, p = 1, c = 0, t = 10.66
sf = 0.5, d = 16, p = 1, c = 0, t = 12.1
sf = 1, d = 16, p = 1, c = 0, t = NaN

References

Brainard, DH (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.

Banks MS, Gepshtein S, and Landy MS. (2004), Why is Spatial Stereoresolution so Low? *J. Neuro*, 24 (9), 2077-2089.

Campbell FW and Robson JG (1968) Application of Fourier Analysis to the Visibility of Gratings. J. Physiol., (197), 551-566.

Collett TS, Schwarz U, Sobel EC. (1991) The interaction of oculomotor cues and stimulus size in stereoscopic death constancy. *Perception,* 20(6):733-54.

Didyk P, Ritschel T, Eisemann E, Myszkowski K, and Seidel HP (2011). A perceptual model for disparity. *ACM Trans. Graph.,* 30, 4, Article 96

Dodgson NA (2004). Variation and extrema of human interpupillary distance. *SPIE: Stereoscopic Displays and Applications,* 5291, 36–46.

Emmert E. (1881) Größenverhältnisse der Nachbilder. Klinische Monatsblätter für Augenheilkunde und für augenärztliche Fortbildung, 19: 443-450.

Ernst MO and Banks MS. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429-33.

Fleet DJ, Wagner H, Heeger DJ. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res.*, 36(12), 1839-57.

Filippini HR and Banks MS. (2009). Limits of stereopsis explained by local cross-correlation. *J. Vis.*, 9 (1), 8.1-818.

Fründ, I, Haenel, NV, Wichmann, FA. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *J. Vision*, 11 (16).

Georgeson MA, and Sullivan GD (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *J. of Phys.,* 252(3), 627-656.

Held RT and Banks MS (2008). Misperceptions in stereoscopic displays: a vision science perspective. *Applied perception in graphics and visualization (APGV '08)*. ACM, 23-32.

Hoffman DM, Girshick AR, Akeley K, Banks MS (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *J. of Vision*, 8(3), 33.

Kane D, Guan P, Banks MS. (2014). The Limits of Human Stereopsis in Space and Time. *J. Neuro*, 34 (4), 1397-1408.

Lankheet, MJM and Lennie P (1996). Spatio-temporal requirements for binocular correlation in stereopsis. Vis. Res. 36, 527-538.

Liang J, Grimm B, Goelz S, and Bille JF (1994). Objective measurement of wave aberrations of the human eye with the use of a Hartmann–Shack wave-front sensor. *J. Opt. Soc. Am. A*, 11, 1949-1957

Love GD, Hoffman DM, Hands PJW, Gao J, Kirby AK, Banks MS (2009). High-speed switchable lens enables the development of a volumetric stereoscopic display. *Opt. Express,* 17(18), 15716–15725

Nienborg, H, Bridge, H, Parker, AJ and Cumming, BG. (2004) Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. J. Neurosci. 24: 2065-2076.

Maloney LT and Wandell BA (1986). Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A* 3, 29-33.

"Perceptual Depth Compression for Stereo Applications"

Pająk D, Herzog R, Mantiuk R, Didyk P, Eisemann E, Myszkowski K, Pulli K (2014). Perceptual Depth Compression for Stereo Applications. *Proc. of Eurographics 2014.*

Pelli, D.G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10:437-442.

Qian N and Zhu Y (1997). Physiological computation of binocular disparity. *Vis Res.,* 37(13): 1811–1827

Rogers BJ, Bradshaw MF (1993). Vertical disparities, differential perspective and binocular stereopsis. Nature, 361, 253–255.

Rogers BJ, Bradshaw MF (1995). Disparity scaling and the perception of frontoparallel surfaces. Perception, 24, 155–179.

Santamaría J, Artal P, and Bescós J (1987). Determination of the point-spread function of human eyes using a hybrid optical–digital method. *J. Opt. Soc. Am. A*, 4, 1109-1114

Shibata T, Kim J, Hoffman D, Banks MS (2011). The zone of comfort: Predicting visual discomfort with stero displays. *Journal of Vision,* 11(8).

Tyler CW. (1974). Depth Perception in Disparity Gratings. *Nature*, 251, 140–142.

Wallach H and Zuckerman C (1963). The Constancy of Stereoscopic Depth. *The American Journal of Psychology,* 76(3), 404-412

Watt SJ, Akeley K, Ernst MO, Banks MS (2005). Focus cues affect perceived depth. *Journal of Vision,* 5(10): 834-862

Woods, A. J., Docherty, T., and Koch, R. 1993. Image distortions in stereoscopic video systems. *Proc. of the SPIE,* 1915, Stereoscopic Displays and Applications IV, J. O. Merritt J. O. and S. S. Fisher, Eds., 36--47.

Chapter 2: Disparity and Luminance Cue Integration in the Perception of Edge Sharpness

Introduction

The visual system combines a multitude of visual cues to synthesize the world we perceive, but it is not equally sensitive to all of these cues. Contrast gratings, for example, can be perceived up to ~60 cycles per degree (cpd) in luminance (Campbell and Robson 1968), but an equivalent grating can only be seen at spatial frequencies up to approximately 3cpd in disparity (Tyler 1974, Banks et al. 2004) and 10cpd in chrominance (Mullen 1985). This means that humans can detect much finer variations in luminance compared to disparity or color. Image and video compression algorithms take advantage of this fact and allocate mor bits for luminance information using than chrominance information (Wallace 19e91) because the visual system is able to combine the luminance information with the color information to create the appearance of sharp color boundaries.

Similar to color, depth edges that occur at occlusion boundaries appear as sharp as edges that define luminance boundaries. Knowing that edges that appear sharp in disparity (Kane et al. 2014) would appear quite blurred in luminance (Watt and Morgan 1983, Hess et al. 1989) we wondered how the visual system integrates luminance cues with disparity cues to dictate the appearance of depth edges. Do depth edges appear sharp because they are coincident with a sharp change in luminance and would a sharp depth discontinuity appear gradual if it was coincident with a gradual change in luminance?

Using luminance information to "fill-in" for disparity information is only sensible if depth edges are reliably coincident with luminance edges in the natural environment. The physical attributes responsible for changes in luminance and disparity gradients are different; luminance gradients in the world can be attributed to three main factors: changes in albedo, cast shadows, and highlights whereas disparity gradients result from changes in the structure of an object and occlusion boundaries. Studies investigating natural scene statistics and have found statistical correlations between luminance edges and depth edges (Fowlkes et al. 2007, Liu et al. 2009, Yang et al. 2010, Vilankar et al. 2014). Furthermore, there is also evidence that the visual system utilizes these scene statistics to in depth perception (Simoncelli and Olshausen 2001, Geisler 2009, Burge et al. 2010, Cooper and Norcia 2014).

Edges can be defined using multiple cues and it is important to understand how those cues are integrated. Landy and Kojima (2001) examined texture edges defined by contrast and orientation or spatial frequency and orientation and found that the contribution of the various cues to the perception of the edges could be explained by weighting the contribution of each cue according to its reliability in a manner consistent with optimal cue combination (Knill and Saunders 2003, Ernst and Banks 2002, Hillis et al. 2004). We presented smoothed steps in luminance and in disparity in order to determine the contribution of luminance and disparity to the perception of an edge. We put the luminance- and disparity-defined edges in conflict with each to investigate the cue combination strategy used by the visual system to integrate them.

General Methods

Apparatus. The experiments were conducted on a Macbook Pro 13" Retina Display using a mirror haploscope to provide a vergence distance that matched the physical screen distance of 128cm. Display resolution was 2560x1600 and refresh rate was 60Hz. Each pixel subtended 0.3arcmin at the viewer's eyes. Maximum and minimum luminances of the stimuli were respectively 266 and 0.4cd/m^2. Gamma correction was applied. The experiment was conducted in a dark room so that the display screen was the only visible input to the eyes.

Observers. Four observers (one male, three female) 21-25 years of age participated in all of the experiments. All had normal or corrected-to-normal vision. Their stereo vision was checked using a standard clinical test. One observer was an author; the other three were naïve to the purpose of the experiments.

Experiment 1: Single-Cue Experiment Methods

Prior to investigating the interaction of luminance and disparity we first measured blur-detection thresholds for luminance-defined edges and for disparity-defined edges to determine the blur values that could be usefully employed in the two-cue experiment. We also measured the just-noticeable differences at suprathreshold levels of luminance and disparity blur to obtain cue combination predictions for the weights of each cue in the two-cue experiments.

Luminance Blur Detection. The luminance stimulus was a rectangle that was 2.4° high and 4.3° wide. The rectangle's height and vertical position were randomly adjusted trial by trial by up to 0.15°. The interior of the rectangle was brighter than the surround; we refer to the luminances of the interior/surround regions as L_{max} and L_{min}, respectively. The top and bottom edges of the rectangle differed in how sharply the luminance changed from L_{max} to L_{min}; both edges of the rectangle were created by convolving two step functions (with values L_{max} and L_{min}) with two Gaussians having standard deviations of either 0 (i.e. a delta function) or σ_L. In the sharp edge the luminance changed from L_{max} in one pixel row to L_{min} in the adjacent row. In the blurred edge, the luminance changed over more rows. Stimuli were rendered in MATLAB with the PsychToolbox (Brainard 1997, Pelli 1997) and presented for 1sec. A uniform gray field with a luminance of 94cd/m^2 was presented between stimulus presentations. A fixation target composed of two binocular horizontal line segments and two dichoptic vertical line segments was presented all the time. By monitoring the apparent alignment of the dichoptic segments, observers could make sure that fixation was accurate before and during stimulus presentations. The observer's task was to indicate which of the two edges was blurred. No feedback about the correctness of the response was provided. 220 to 750 trials were presented to each observer for each condition. The experiment was conducted with two luminance contrasts: 0.71 and 0.13, where contrast is defined as $(L_{max} - L_{min})/(L_{max} + L_{min})$.

We varied σ_L trial by trial using the method of constant stimuli. We fit the resulting psychometric data with a cumulative Gaussian using a maximum-likelihood criterion (Fründ et al. 2011). The mean of that fitted function was the estimate of the value of σ_L for which observers could correctly identify the blurred edge on 75% of the trials.

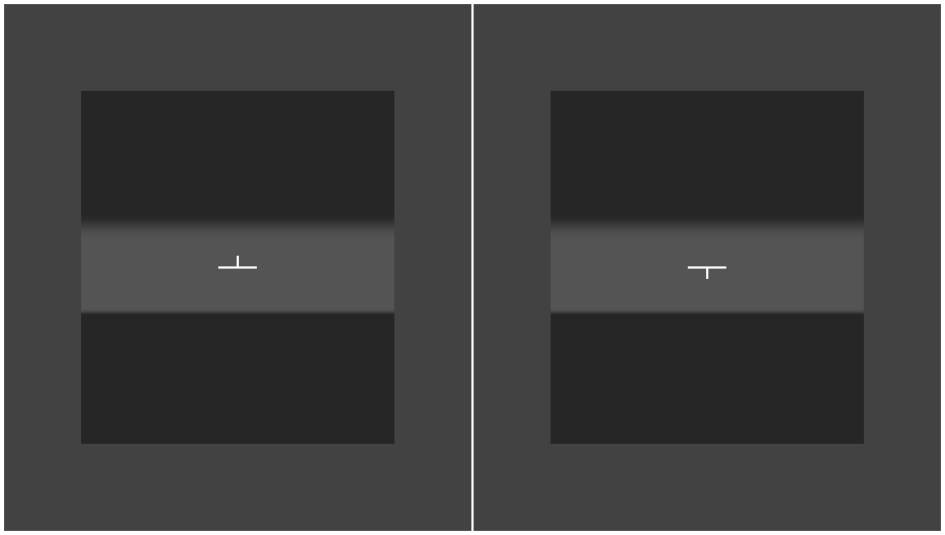


Figure 23 Luminance stimulus. The rectangular stimulus had a bright interior with a luminance of L_{max} and a dark surround with a luminance of L_{min}. One horizontal edge (either the one at the top of bottom of the rectangle) had a sharp change in luminance while the other had a gradual change. Here the top edge of the rectangle is blurred and the bottom edge is sharp. A dichoptic fixation target assured that the observer was converged accurately on the stimulus before each presentation.

Disparity Blur Detection. We used an analogous procedure to also measure the ability to detect blur in a disparity-defined edge. In this case, a random-dot stereogram depicted a rectangle raised in depth. Dot density was 144 dots/degree^2 yielding a per-frame Nyquist limit of 6cpd (Banks, Gepshtein, & Landy, 2004). Dots were refreshed at 60Hz, so the effective Nyquist limit was much higher (Lankheet & Lennie, 1996). Dots were 0.9arcmin in diameter, 266cd/m^2 bright, and drawn using anti-aliasing to create sub-pixel displacements. The central rectangular region containing depth had a crossed disparity of 8arcmin and was 2.3° tall and 4.3° wide with ±0.15° jitter in height and vertical position. The surround had a disparity of 0arcmin. The disparity of the sharp edge of the rectangle changed from eight to 0arcmin in adjacent pixel rows. The disparity at the blurred edge changed smoothly from eight to 0arcmin at the rate determined via convolution with a Gaussian-blur kernel with standard deviation σ_D. There were unmatched dots (i.e., seen by one eye, but not the other) near the left and right edges of the stimulus, but they were not noticeable because they were far from fixation.

On each trial, observers indicated which of the two edges appeared blurred. We used the method of constant stimuli to vary σ_D trial by trial. 360 to 660 trials were presented per observer and condition. No trial-by-trial feedback was provided. We fit a cumulative Gaussian to the resulting psychometric data (Fründ et al. 2011). The mean of the fitted function served as the estimate of the value of σ_D that yielded 75% correct response.

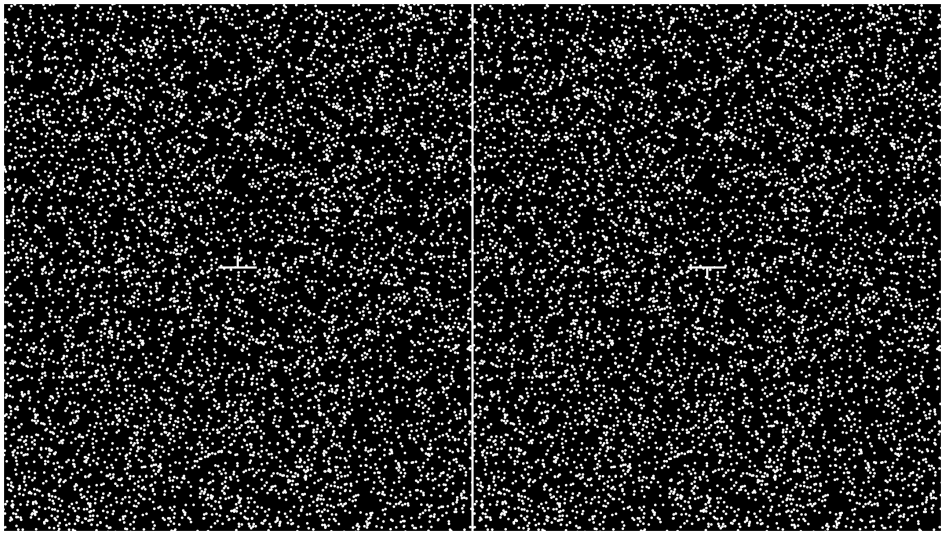


Figure 24 Disparity RDS stimulus. The rectangular stimulus had a raised interior with a disparity of 8arcmin and a uniform surround with a disparity of 0arcmin. One horizontal edge (either the one at the top of bottom of the rectangle) had a sharp change in disparity while the other had a gradual change. Here the top of the rectangle is blurred and the bottom edge of the rectangle is sharp. A dichoptic fixation target assured that the observer was converged accurately on the stimulus before each presentation.

Blur Discrimination. Having established the blur values that were suprathreshold, we turned to measuring single-cue, blur-discrimination thresholds for luminance-defined and disparity-defined edges. In this experiment, one edge had a reference blur value (σ_L or σ_D = 3.2, 4.8, or 6.8 arcmin) and the other had a comparison value σ that was varied from trial to trial. Observers indicated on each trial which of the two edges appeared blurrier. The blur value of the comparison stimulus was varied according to the method of constant stimuli. We fit the resulting psychometric data with a cumulative Gaussian ranging from 0 to 100% correct. Here the point of subjective equality (PSE) is not very informative because subjects are expected to be 50% correct should when $\sigma = \sigma_L$ or σ_D. Instead, we were interested in measuring the just-noticeable differences (JNDs) for each value of σ_L or σ_D. We calculated the JND for each condition by determining the values of σ_{25} and σ_{75} corresponding to 25% and 75% percent correct in the fitted psychometric function and taking their difference and diving by two (JND = $(\sigma_{75}-\sigma_{25})/2$).

Experiment 1: Single-Cue Experiment Results

Luminance and Disparity Detection Thresholds. At 71% contrast the smallest detectable σ_L was 0.55arcmin which is in agreement with previous studies (Hess et al. 1989, Watt and Morgan 1983). The smallest detectable σ_D was 1.25arcmin which also agrees with previous studies (Kane et al. 2014). Figure 25 illustrates the differences in these sensitivities and the thresholds are also shown in Figure 28.

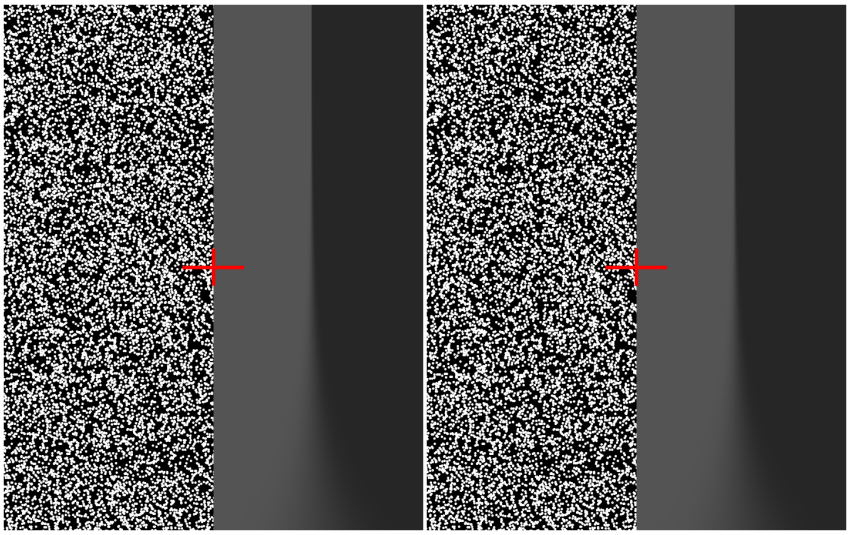


Figure 25 Cross-fuse to view a blurred edge in luminance and disparity presented side-by-side. The blur of each edge is smallest at the top of the image and largest at the bottom. The magnitude of the blur is the same for cues. A smaller amount of blur can be detected in the luminance edge (blur is identified closer to the top of the image) compared to the disparity edge illustrating the differences in blur sensitivity.

Luminance and Disparity JNDs. Figure 26 shows the blur-discrimination thresholds for the high- and low-contrast luminance-defined edges (upper and middle panels) and for the disparity-defined edges (lower panel). Thresholds were quite similar across observers. Not surprisingly, thresholds for detecting or discriminating changes in luminance-defined edges were lower when the edges had high luminance contrast; i.e., observers were able to perceive small changes in blur with high-contrast edges. Thresholds were consistently higher with disparity-defined edges than with luminance-defined edges, meaning that observers were less able to perceive small blur changes when the edge was defined by disparity. Thus, as we expected, observers are less able to detect a spatial change in disparity than a corresponding change in luminance.

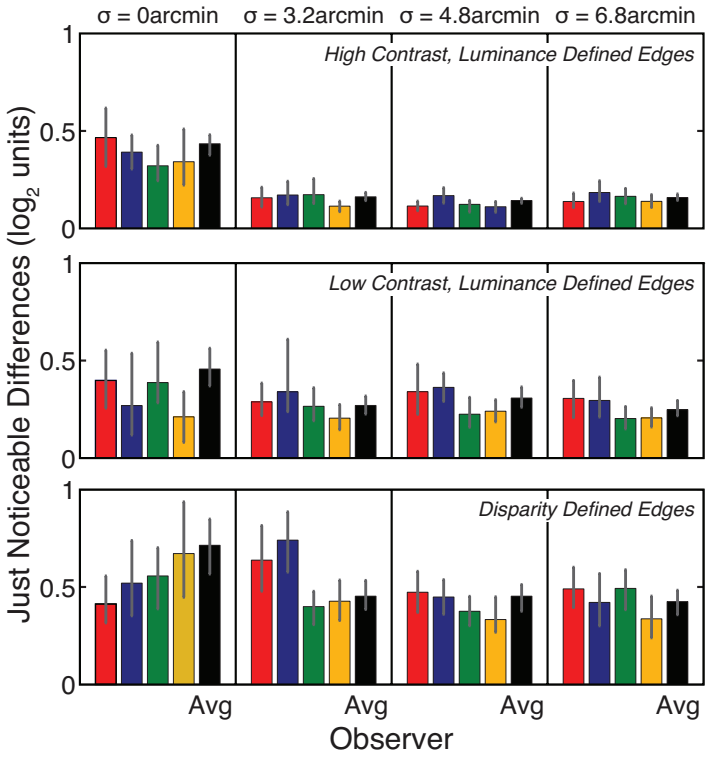


Figure 26 Just-noticeable differences for high contrast, low contrast, and disparity edges. The JNDs from each psychometric function is plotted for each condition and observer. Each column represents the value of σ used to create the blurred edge. The four colored bars in each panel represent thresholds for individual subjects; the black bars represent thresholds when the data are combined across subjects. The upper, middle, and lower panels represent the data respectively for high-contrast, luminance-defined edges, low-contrast, luminance-defined edges, and disparity-defined edges. Error bars represent 95% confidence intervals.

Experiment 2: Two-Cue Experiment Methods

We next examined how luminance and disparity information was combined when estimating the blur of an edge. To do so, we presented random-dot stimuli with variation in luminance and disparity.

Stimulus and Methods. In all stimuli, one edge had the same luminance- and disparity-defined blur (cues consistent, $\sigma_D = \sigma_L$) and the other edge had different luminance- and disparity-defined blurs (cues inconsistent, $\sigma_D \neq \sigma_L$). An example stimulus is shown in Figure 27 along with its corresponding luminance and disparity profiles. Observers indicated which of the two edges appeared blurrier. Dot density was again 144 dots/deg^2. Each dot had a diameter of 0.9arcmin. The interior of the rectangle was brighter than the surround. The interior and surround contained randomly positioned dots such that that interior had crossed disparity of 8arcmin and the surround had a disparity of 0arcmin. The luminances of the dots were adjusted such that the Weber contrast of the dots relative to background was always constant and equal to 12.

As before, a fixation target composed of two binocular and dichoptic segments was presented at all times. Observers were told to maintain fixation on the target throughout. Observers indicated after each stimulus presentation whether the top or bottom edge appeared blurrier. No feedback was provided.

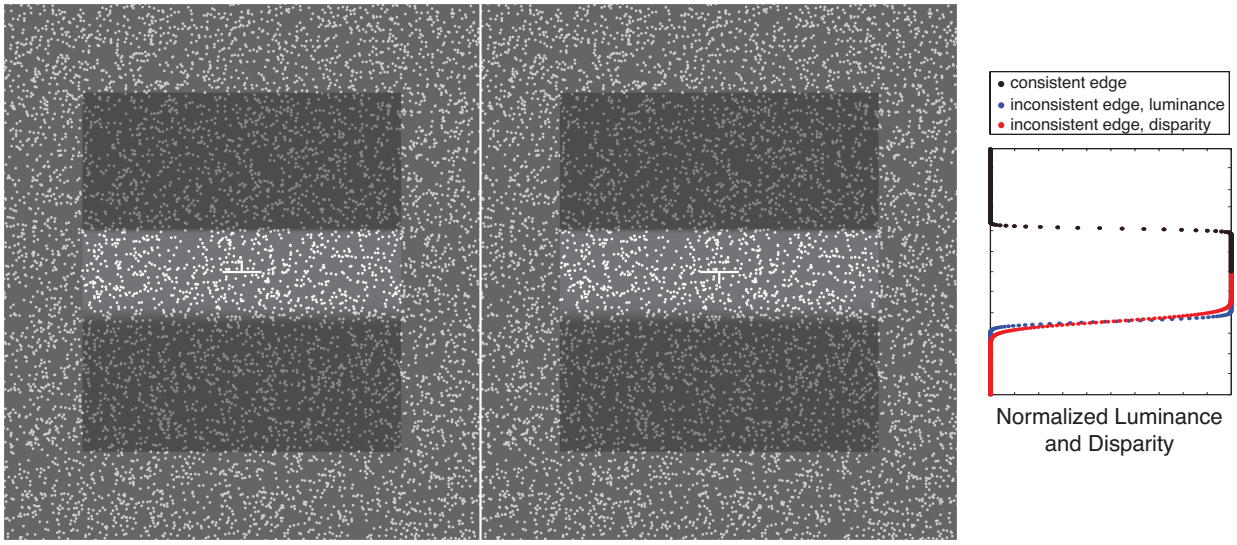


Figure 27 Two-cue stimulus. The interior of the rectangle was brighter than the surround. The interior also had crossed disparity relative to the surround. This created the appearance of a raised rectangle brighter than the surround. One horizontal edge of the rectangle had the same blur kernels applied to the luminance and disparity ($\sigma_D = \sigma_L$); this was the cues-consistent stimulus. The other horizontal edge had different kernels applied to the luminance and disparity ($\sigma_D \neq \sigma_L$); this was the cues-inconsistent stimulus. The profiles on the right show the different gradients for the cues-consistent (upper) and cues-inconsistent (lower edges).

Figure 28 schematizes the experimental conditions for the two-cue experiment. Each stimulus contained a cues-inconsistent and a cues-consistent edge. There were 16 cues-inconsistent stimuli, four at each of four base blur values; they are represented by the red and green circles. For each of the cues-inconsistent stimuli, we sought the blur value of the cues-consistent stimulus that made the edge gradients appear to be the same. To find this value, we used the method of constant stimuli to vary the blur of the consistent edge. As in the single-cue experiment, two luminance contrasts were used. In the high-contrast condition (contrast = 0.71) all four base blur values (0.6, 1.2, 2.4, and 4.8arcmin) were tested for a total of 16 conditions. In the low-contrast condition (contrast = 0.13), only one base blur value (4.8arcmin) was tested. 300-530 trials were presented in each condition for each observer.

We fit the psychometric data obtained in each condition for every observer with cumulative Gaussians. The points of subjective equality (PSEs) were the 50% points on the fitted functions. We defined the just-noticeable differences (JNDs) as the change in blur value between the 25% and 75% points on the fitted functions divided by two.

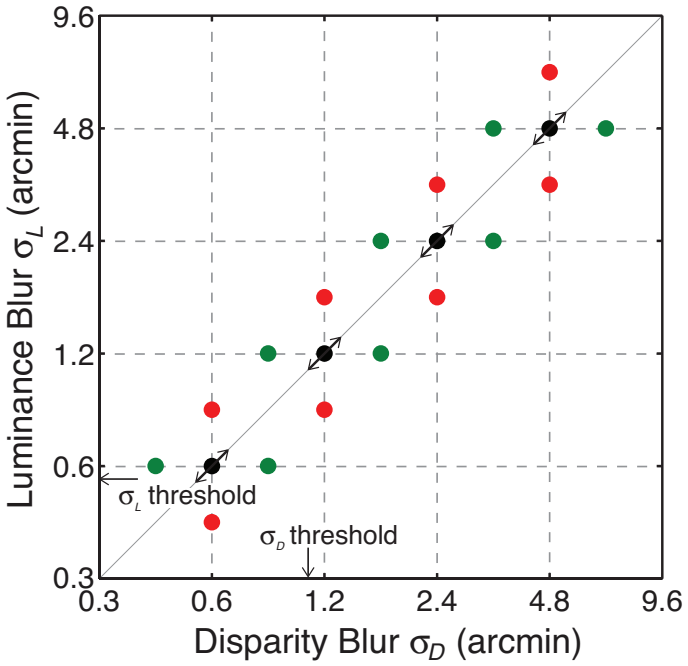


Figure 28 Two-cue experimental conditions. The standard deviation of the blur of the disparity-defined edge is represented on the abscissa and standard deviation of the blur of the luminance-defined edge on the ordinate. The gray diagonal line represents cues-consistent stimuli in which the blurs of the luminance-defined and disparity-defined edges are the same. We examined four blur levels: 0.6, 1.2, 2.4, and 4.8arcmin, each indicated by five circles in a grouping. The green points represent conditions in which the disparity blur kernel was perturbed from the central blur level and the red points represent conditions in which the luminance blur kernel was perturbed. The cues-consistent stimuli are represented by the black circles; the arrows indicate that the blur value of those stimuli was changed until they had had the same apparent blur as the cues-inconsistent stimuli. The arrows near the abscissa and ordinate indicate the blur-detection thresholds.

Experiment 2: Two-Cue Experiment Results

The cues-consistent edge that will appear to have the same gradient as an inconsistent edge depends on the relative weights given to the luminance and disparity cues. Figure 29 shows three sets of predicted PSEs with different relative weights for the luminance and disparity cues. Let w_L and w_D (where $w_L + w_D = 1$) be the weights assigned to luminance and disparity, respectively. The left panel shows when the perceived gradient is determined entirely by the luminance information (w_L =1, w_D = 0). The middle panel shows the predicted PSEs when the perceived gradient is determined equally by the luminance and disparity information ($w_L = w_D = 0.5$). The right panel shows the predictions when the perceived gradient is determined solely by the disparity information (w_L =0, w_D = 1).

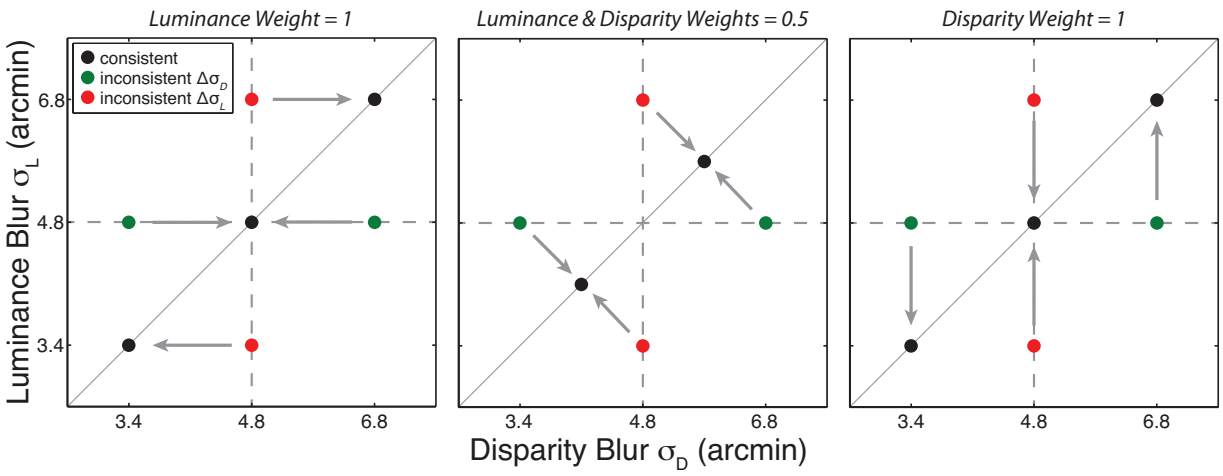


Figure 29 Predicted PSEs for two-cue experiment. In each panel the standard stimuli (cues-inconsistent) are represented by the red and green circles and the comparison stimuli (cues-consistent) are represented by the black circles. The panels from left to right show the predicted PSEs if the luminance weight is 1 (and the disparity weight is 0), if the luminance weight is 0.5 (and disparity weight 0.5), and if the luminance weight is 0 (and disparity weight is 1). The arrows indicate the value of the reference stimulus that should be perceived as the same as the comparison stimulus under the three weighting assumptions and for each comparison stimulus.

The results of the high-contrast cue conflict experiments are shown in Figure 30. The cues consistent blur for each conflict is plotted along the cues consistent line with 95% confidence intervals. There are seven conditions where (σ_D <=1.2 arcmin) where σ_D is near the measured sensitivity threshold to disparity blur. For these conflicts it is unsurprising that the perceived sharpness closely matches the value of σ_L. However, even when σ_D was well above threshold at 2.4 and 4.8 arcmin the perceived edge sharpness is still closely matched to σ_L. A notable difference in the reported PSE values for this range of conflicts is that the confidence intervals are much smaller for conditions where both the luminance and disparity cues are above threshold. Figure 31 illustrates how apparent blur is affected by the luminance and disparity cues.

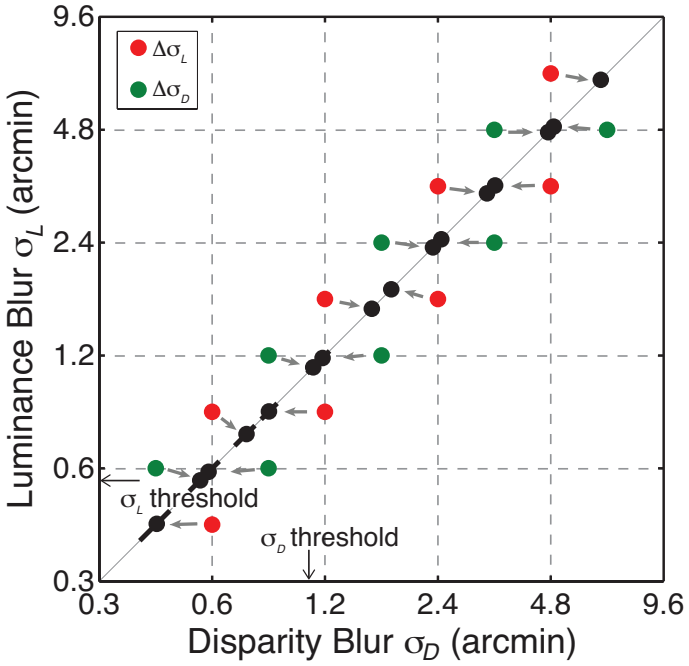


Figure 30 High Two-cue high contrast results. The 16 cue-conflict conditions are represented by the red and green dots and their corresponding cues consistent PSEs are shown in black along the gray cues consistent line. In nearly all 16 conditions the PSE is heavily weighted towards σ_L. For the five conditions where σ_L <= 0.85arcmin the confidence intervals are much larger than the other 10 conditions.

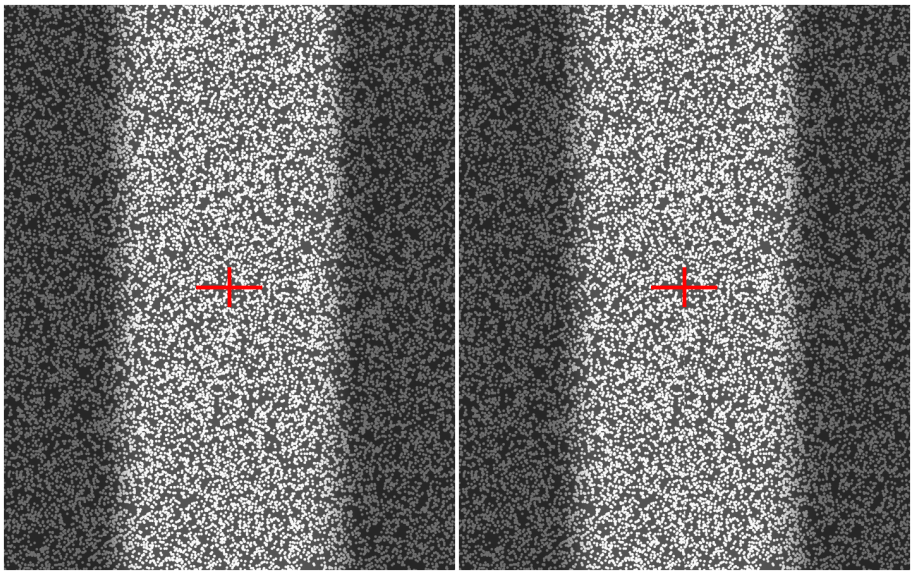


Figure 31 Luminance and disparity cue-conflict edges. Both edges have conflicting amounts of luminance- and disparity-defined blur. The disparity-defined blur for the left edge is minimum at the top of the image and maximum at the bottom of the edge while the luminance-defined blur is constant. In the right edge the luminance-defined blur is smallest at the top of the image and largest at the bottom of the image, the change in blur is equal to the change in disparity blur in the left edge. Likewise, the disparity-defined blur is constant and equal to the luminance-defined blur in the left edge. The left edge should have the appearance of a constant blur regardless of elevation while the apparent blur of the right edge should increase near the bottom of the image.

Disparity and Luminance Cue Integration

The visual system has been shown to use maximum likelihood estimation to resolve conflicts between multiple cues (Ernst and Banks 2002). We wondered if the measured PSEs from the two-cue experiment could be predicted using maximum likelihood estimation and the JNDs associated with the single cue experiments for σ_L and σ_D = 3.2, 4.8, and 6.8arcmin from the single cue experiment (Figure 26). We used those JNDs to estimate the weights associated with the luminance and disparity cues to predict perceived sharpness according to maximum likelihood estimation where:

$$w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2} \tag{1}$$

We calculated the w_L and w_D for the four cue conflicts with a base blur of 4.8arcmin and show those results are shown in Figure 32A. The calculated weights for luminance from MLE and the empirical luminance weights from the measured PSEs are shown in Figure 32B.

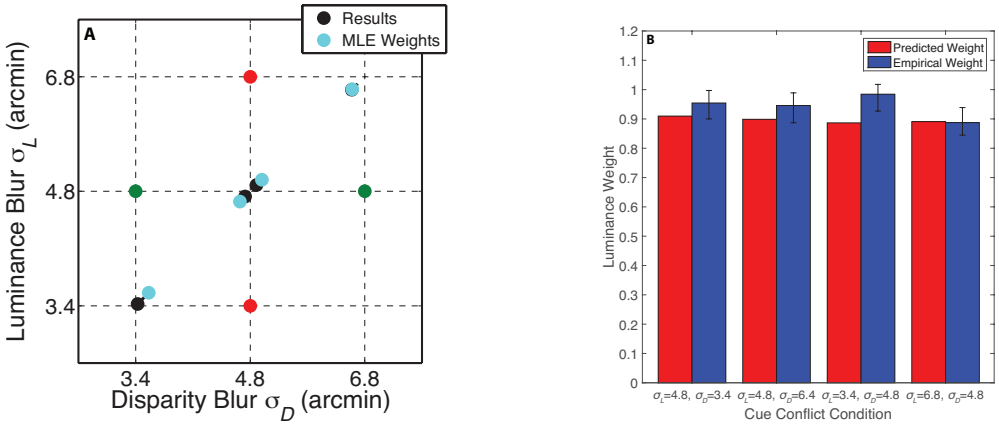


Figure 32 High contrast stimulus with maximum likelihood estimation and weights. **A)** The red points represent conflicts where σ_D is fixed at 4.8arcmin and σ_L is varied, the green points represent conflicts where σ_L is fixed at 4.8arcmin and σ_D is varied. Black circles represent the cues consistent edge ($\sigma_L = \sigma_D$) that appears to have the same amount of blur and cyan circles represent the predicted values using maximum likelihood estimation. **B)** Predicted luminance weights using maximum likelihood estimation compared to empirical luminance weights.

According to (1) cues with a smaller variance will receive greater weight and in a two-cue experiment the cue with the smaller variance will receive a majority of weight if the differences in variance are large. In the high contrast conditions the JNDs of the luminance cue was less than half the JND of the disparity cue and this is reflected in Figure 32B; the luminance cue is weighted by approximately 90% in all four conflicts. Reducing the contrast from 71% to 13% also increases the JND of the luminance signal (Figure 26) which increases the theoretical weight of the disparity cue in the two-cue condition. We repeated the conditions of the two-cue experiment seen in Figure 32 at 13% contrast. As expected, the perceived edge sharpness is more heavily influenced by disparity when the variance of the luminance cue is higher (Figure 33A); in all cases reducing the contrast of the luminance signal increases the contribution of the disparity signal in the perceived

sharpness. The calculated weights for luminance from MLE and the empirical luminance weights from the measured PSEs are shown in Figure 33B; the theoretical luminance weights have fallen from ~90% to between 60-70%.

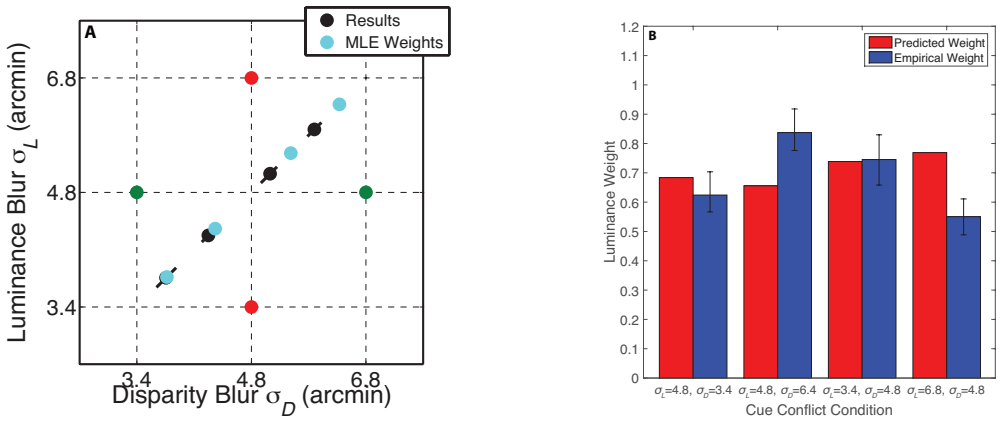


Figure 33 Low contrast stimulus with maximum likelihood estimation and weights. **A)** The red points represent conflicts where σ_D is fixed at 4.8arcmin and σ_L is varied, the green points represent conflicts where σ_L is fixed at 4.8arcmin and σ_D is varied. Black circles represent the cues consistent edge ($\sigma_L = \sigma_D$) that appears to have the same amount of blur and cyan circles represent the predicted values using maximum likelihood estimation. **B)** Predicted luminance weights using maximum likelihood estimation compared to empirical luminance weights.

Maximum likelihood estimation also predicts that the variance in the two-cue stimulus should be smaller than the variance of any single cue (in this case the luminance-only conditions). We calculated the JNDs associated with each of the two-cue psychometric functions and compared them to the predictions generated by MLE. Contrary to the MLE prediction. As seen in Figure 34 we found that JNDs increased in the two-stimulus conditions relative to the luminance-only JNDs. In both the low- and high-contrast conditions the empirical JNDs were much larger than the predicted JNDs from an optimal cue combination model.

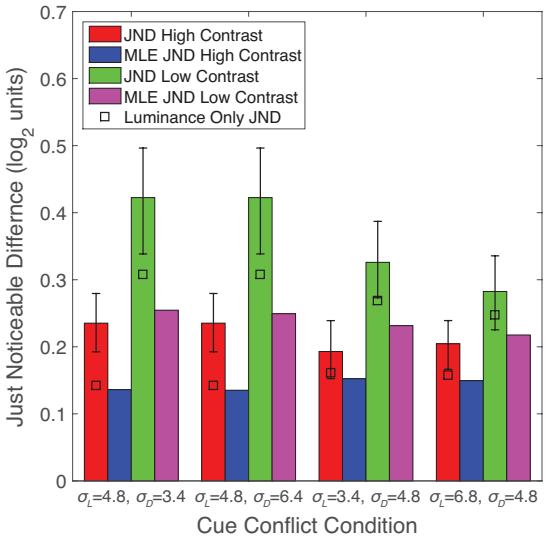


Figure 34 Empirical JNDs vs. predicted JNDs from MLE. We show the predicted and empirical JNDs for both contrast conditions at all four cue conflicts. Red and green bars represent empirical JNDs for the

high and low contrast conditions respectively. Blue and magenta bars represent the predicted JNDs for high and low contrast conditions. In optical cue combination the JNDs in a multi-cue stimulus should be less than the smallest single-cue JND (which is the luminance only JNDs represented by the black squares). We found that the empirical JNDs increased in the two-cue stimulus conditions compared to the single-cue conditions suggesting that the visual system is not using optimal cue combination in our task.

One alternative to optimal cue combination is that subjects employ a cue-switching strategy (Landy and Kojima 2001, Oruc et al. 2003). In cue switching the predicted weights from MLE would indicate the proportion of times each cue was used to make a judgement. We used the weights obtained from maximum likelihood estimation to calculate the predicted cue-switching psychometric functions by taking the weighted sum of the single-cue disparity and luminance psychometric functions. Figure 35 shows the theoretical cue-switching psychometric functions with the fitted psychometric functions obtained from our data. Due to the extremely high luminance weighting in the high contrast conditions, the cue-switching prediction and observed psychometric function are both similar to the high contrast luminance-only psychometric function. In the low contrast conditions the disparity cue has a larger weight and there is more separation between the observed psychometric function and the luminance only cue. The cue switching predictions appear to provide a good fit to the data in this case as well.

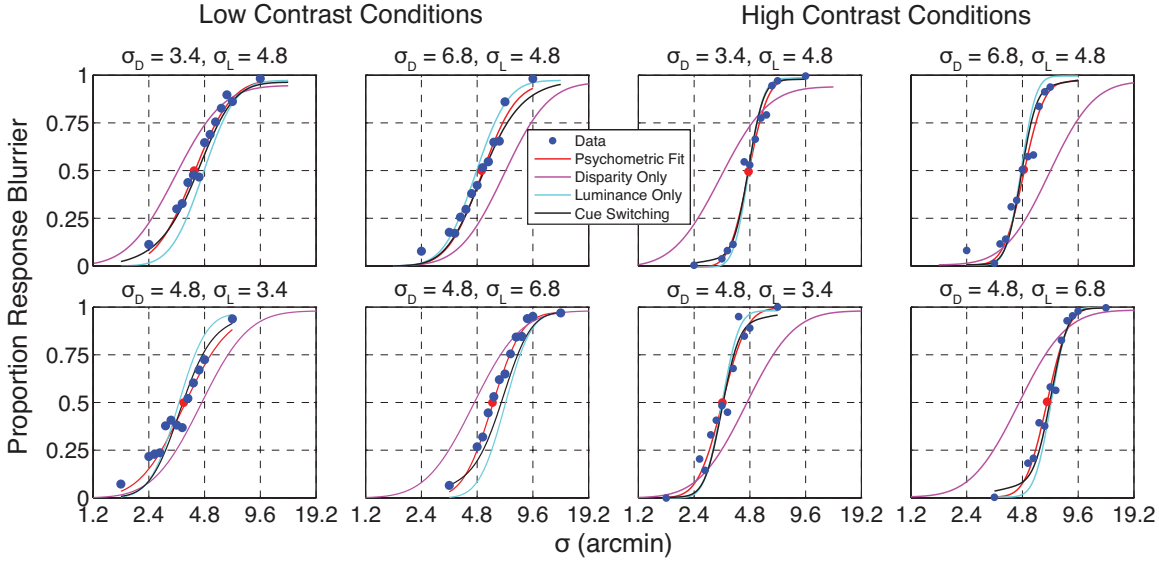


Figure 35 Predicted psychometric functions obtained using cue switching compared to empirical psychometric functions. The left four panels refer to the low contrast two cue conditions and the right four panels refer to the high contrast two cue conditions. In all panels the purple and cyan lines represent the single-cue luminance and disparity psychometric functions; the blue circles and red lines represent the data and empirical psychometric fit. The cue switching prediction as created by weighting the luminance and disparity only psychometric functions according to Equation 1. Using the empirical weights obtained for luminance and disparity yields even better matches between the cue switching psychometric functions and the empirical fits.

We also calculated the predicted JNDs from the theoretical cue switching psychometric functions and show those predictions in Figure 36. Unlike MLE, cue switching predicts that the JNDs will increase relative to the cue with the smallest variance. In the high contrast conditions the cue switching psychometric function is heavily weight so there the predicted JNDs are only slightly larger than the JNDs derived from the single-cue

luminance only psychometric functions. In the low contrast conditions luminance weight decreases substantially and the predicted JNDs for the two-cue conditions also rise significantly. In fact the JNDs predicted by cue-switching agree with the empirical JNDs from our data providing support to the claim that the visual system is using cue switching and not optimal cue combination in our task.

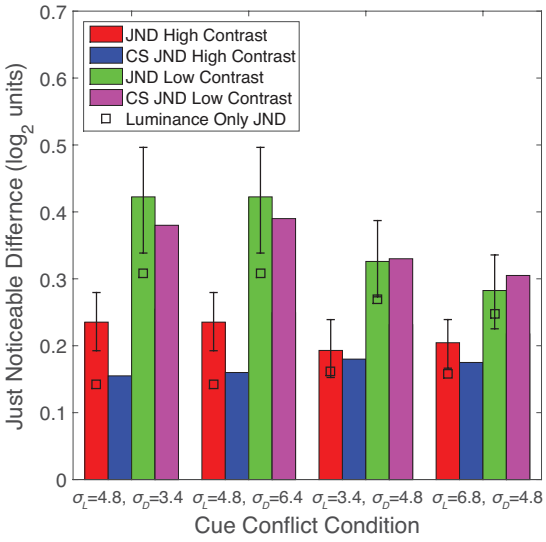


Figure 36 Empirical JNDs vs. predicted JNDs from cue switching (CS). We show the predicted and empirical JNDs for both contrast conditions at all four cue conflicts. Red and green bars represent empirical JNDs for the high and low contrast conditions respectively. Blue and magenta bars represent the predicted JNDs for high and low contrast conditions. In cue switching the JNDs in a multi-cue stimulus should be larger than the smallest single-cue JND (which is the luminance only JNDs represented by the black squares). We found that the empirical JNDs increased in the two-cue stimulus conditions compared to the single-cue conditions and the predicted JNDs match the empirical JNDs.

Discussion

Appearance of Luminance- and Disparity-Defined Edges

Robinson and MacLeod (2013) investigated the apparent position of an edge defined by both luminance and disparity. In their experiment the two edges were either co-incident, separated by 2.8arcmin, or separated 5.6arcmin. When the edges were separated, they found that the apparent location of a two-cue edge was more heavily weighted towards the luminance edge. Using variances obtained from single-cue measurements, Robinson and MacLeod concluded that optimal cue combination predicted some, but not all of their effect. They also note that the variances of their two-cue stimuli were larger than predicted by MLE. However, they rejected cue switching as a plausible alternative in their data because in their cue-switching modeling at 5.6arcmin separation the simulation showed two inflection points in the predicted psychometric function while they only observed a single inflection point in their two-cue psychometric functions. In our stimulus the cue conflict between σ_L and σ_D were deliberately chosen to ensure that the cue-conflict edge appeared as a single edge rather than two distinct edges visible separately in luminance and disparity. Robinson and MacLeod (2013) claim that large conflicts should result in psychometric functions with multiple inflection points. If we had used a larger conflict it is likely that there would be multiple inflection points in the cue-switching psychometric function, but that would likely be attributed to the perception of two distinct

edges. Regardless of the cue combination rule, data from Robinson and MacLeod and our experiments suggest that spatial structure of edges (location and sharpness) are primarily dictated by luminance; a fact which could be relevant in stereo-compression algorithms (Didyk et al. 2012, Pająk et al. 2014).

Blur and Disparity Gradients at Occlusion Boundaries

There is an upper bound to the largest disparities that can support the sensation of depth; this limit is a result of the local cross-correlation used to calculate disparity and is called the disparity-gradient limit (Banks et al. 2004, Kane et al. 2014). The disparity gradient at an occlusion edge is infinite which means that the visual system cannot use disparity to estimate the depth between two occluding surfaces at the exact location of the occlusion. Blur and disparity are complimentary depth cues (Held et al. 2012) and the blur gradient at an occlusion edge can be used to estimate absolute depth at the occlusion edge. This means that the blur or luminance gradient is a more useful depth cue at sharp depth discontinuities. Our results indicate that the luminance cue is generally more reliable and dictates the appearance of depth edges, this is desirable given the viewing geometry at occlusion boundaries. The use of cue switching rather than optimal cue combination observed in our results could be related to the discrepancies between our stimulus and a more realistic occlusion edge and the size of our conflicts (Gepshtein and Banks 2003). The textures associated with both depths at an occlusion boundary are likely unique to each other. Additionally it is rare for each side of the occlusion to differ so drastically in luminance. Finally the vergence and accommodation cues present in our stimulus were not consistent with an occlusion boundary either. Perhaps reducing the size of the conflicts and using more realistic stimuli would result in optimal cue combination.

When does an Edge Appear Sharp?

Figure 25 shows an edge in luminance and an edge in disparity, this demo illustrates that blur in luminance can be detected at smaller amounts than blur in disparity (blur is observed higher in the image). Interestingly, when the luminance edge appears to be blurred, the disparity edge appears to be sharp up until blur is detected. In other words, a perceivable amount of blur in luminance appears sharp in disparity as long as that blur is below the disparity blur sensitivity threshold. Galvin et al. (1997) observed a similar phenomenon in luminance when they compared the apparent sharpness of luminance edges in the fovea and the periphery. In their experiment eleven edges with various amounts of blur were presented at five eccentricities and matched in apparent sharpness to an edge present on the fovea. In all cases the edge presented at the fovea had less blur than the edge in the periphery when the apparent sharpness of the two edges was matched. This suggests that the perception of blur is tied to the blur detection threshold for a particular cue rather than a specific value. Perceptually this means that everything appears sharp until the visual system can detect otherwise. This behavior ensures that the world appears sharp regardless of where the fixation point is; it would be undesirable if an area of the visual field appeared sharp when fixated and then appeared to be blurry immediately after looking away. Our two-stimulus results indicate that the apparent sharpness is combined across cues and that apparent sharpness is determined predominantly by cues that are more reliable.

Conclusions

Depth discontinuities often appear sharp even though visual acuity is poor in the disparity domain. One potential explanation behind this phenomenon is that the visual system uses complimentary information from luminance to determine the location and gradient of depth edges. We investigated this possibility by creating edges with different amounts of blur as specified by luminance and disparity and found that luminance is the primary determinant in perceived edge sharpness. We reduced the contrast of the luminance cue and found that the weight of the luminance cue decreased in the resulting percept in a manner consistent with optimal cue combination. Using single cue psychometric functions we found that the perceived sharpness of the two-cues edge in both the high and low contrast experiments could not be predicted by optimal cue combination and was instead more consistent with a cue switching mechanism. It is possible that using more realistic stimuli with accurately modeled lighting in our representation of a luminance edge or reducing the size of the cue conflicts could result in an optimal luminance and disparity cue combination.

References

Brainard, DH (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.

Burge J, Fowlkes CC, and Banks MS. (2010) Natural-Scene Statistics Predict How the Figure–Ground Cue of Convexity Affects Human Depth Perception. *J Neuro.*, 30(21): 7269-7280; doi: 10.1523

Banks MS, Gepshtein S, and Landy MS. (2004), What is Spatial Stereoresolution So Low? *J Neuro.*, 24 (9), 2077-2089.

Campbell FW and Robson JG. (1968) Application of Fourier Analysis to the Visibility of Gratings. *J. Physiol.,* (197), 551-566.

Cooper EA and Norcia AM. (2014). Perceived Depth in Natural Images Reflects Encoding of Low-Level Luminance Statistics. *J. Neuroscience,* 34(35): 11761-11768

Didyk P, Ritschel T, Eisemann E, Myszkowski K, Seidel HP, Matusik W. (2012). A Luminance-Contrast-Aware Disparity Model and Applications. *ACM Transactions on Graphics*, 31(6)

Ernst MO and Banks MS. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429-33.

Filippini HR and Banks MS. (2009). Limits of stereopsis explained by local cross-correlation. *J Vision.*, 9 (1), 8.1-818.

Fowlkes CC, Martin DR, and Malik J. (2007). Local Figure-Ground Cues are Valid for Natural Images. *Journal of Vision,* 7(8):2, 1-9.

Fründ, I, Haenel, NV, Wichmann, FA. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *J Vision*, 11 (16).

Galvin SJ, O'Shea RP, Squire AM, Govan DG (1997). Sharpness overconstancy in peripheral vision. *Vision Research*, 37(15):2035-2039.

Geisler WS (2008) Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol* 59:167–192.

Gepshtein S, Banks M.S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology,*13:483–488.

Hillis JM, Watt SJ, Landy MS, and Banks MS (2004). Slant from texture and disparity cues: optimal cue combination. *J Vis*, 4:967-992.

Held, R. T., Cooper, E. A., & Banks, M. S. (2012). Blur and Disparity are Complementary Cues to Depth. Current Biology, 22(5), 426–431.

Hess RF, Pointer JS, Watt RJ. (1989). How are spatial filters used in fovea and parafovea? *J. Optical Society of America A, 6,* 329.

Kane D, Guan P, Banks MS. (2014). The Limits of Human Stereopsis in Space and Time. *J. Neuro.*, 34 (4), 1397-1408.

Knill DC and Saunders JA (2003). Do humans optimally integrate stereo and texture information to slant? *Vision Res.,* 43:2539-2558.

Landy MS and Kojima H. (2001). Ideal cue combination for localizing texture-defined edges. *J. Opt. Soc. Am. A,* 18, 2307-2320

Lankheet, M.J.M. & Lennie, P. (1996). Spatio-temporal requirements for binocular correlation in stereopsis. Vis. Res. 36, 527-538.

Liu Y, Cormack LK, Bovik AC. (2009). Luminance, disparity, and range statistics in 3D natural scenes. *Proc of the SPIE*, Volume 7240, id. 72401.

Mullen KT. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *J Physiol.*, 359: 381–400.

Oruc I, Maloney LT, and Landy MS (2003). Weighted linear cue combination with possibly correlated error. *Vis Res.,* 43:2451-2468.

Pająk D, Herzog R, Mantiuk R, Didyk P, Eisemann E, Myszkowski K, Pulli K (2014). Perceptual Depth Compression for Stereo Applications. *Proc. of Eurographics 2014.*

Pelli, D.G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10:437-442.

Robinson AE and MacLeod DIA. (2013). Depth and luminance edges attract. *J of Vision*, 2013;13(11):3. doi: 10.1167/13.11.3.

Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. *Annu Rev Neurosci* 24:1193–1216

Tyler CW. (1974). Depth Perception in Disparity Gratings. *Nature*, 251, 140–142.

Vilankar KP, Golden JR, Chandler DM, Field DJ. (2014). Local edge statistics provide information regarding occlusion and nonocclusion edges in natural scenes. *J Vision,*14(9):13. doi: 10.1167/14.9.13.

Wallace, GK (1991). The JPEG still picture compression standard. *Communications of the ACM,* 34(4), 30-44.

Watt RJ and Morgan MJ (1983). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Ressearch, 23,* 1465-1477.

Yang L, Cormack LK, Bovik AC. (2010). Dichotomy between luminance and disparity features at binocular fixations. *J Vision*. 10(12):23. doi: 10.1167/10.12.23.

Chapter 3 Local Inter-Ocular Contrast Differences are used to Determine Cyclopean Appearance

Introduction

Although the retinal images generated by the left and right eyes are often different from each other due to stereo projection geometry, humans perceive the world in a manner consistent with having a single, cyclopean eye (Wheatstone 1838). Understanding the process of binocular combination is key to understanding how humans perceive the visual scene. Binocular combination has been probed extensively in binocular rivalry literature; however, many of the stimuli used to study rivalry (Tong 2006, Brascamp et al. 2015) present different objects or scenes to the two eyes and are not reflective of the differences that arise from viewing natural images that the visual system has evolved to address (O'Shea 2011, Arnold 2011).

In natural viewing differences that arise in the retinal images from the perspective projection of the world onto the retina at two different vantage points. These differences arising in can be broadly attributed into three separate scenarios: binocular differences in visible reflections; monocular occlusion zones visible to one of the two eyes (Figure 37); and binocular disparity from depth. Monocular reflections and occlusion zones represent cases where one eye has information that the other one does not resulting in local areas of rivalry between the two eyes. Compared to a synthetic stimulus, rivalry in natural viewing occurs primarily at a local scale and can be associated with differences in local image contrast and blur, both of which are important factors in predicating appearance during rivalry (Fahle 1982, Wolfe 1983, Arnold et al. 2007).

Figure 37 Demonstration of local rivalry in natural viewing. In this stereo3D photo the right eye's image (on the left) the "NE" in engineering is occluded by the lamppost and the "ER" is occluded in the left eye's image (on the right). When cross-fused the "N" and "R" are easily visible while the "EE" and lamppost undergo binocular rivalry. In this instance the brain is able to determine which local features are most salient and they are visible in the cyclopean percept.

Figure 37 illustrates the local scale of the rivalry problem in natural vision. Local-cross correlation can successfully explain many unique behaviors observed in disparity detection (Cormack 1991, Fleet 1996, Harris 1997, Read et al. 2002, Banks 2004, Filippini and Banks 2009). We wondered if the binocular combination for appearance might also have similar constraints imposed by the binocular neurons used in local-cross correlations.

In Figure 38 we created stereoimage pairs with regions of monocular blur (i.e. regions of high and low contrast) in different spatial locations of the left- and right-eye images. When the blur is small the cyclopean image appears to be sharp and undistorted because the visual system is able to use the sharp regions from each eye to create a cyclopean image that appears sharp. As the monocular blur regions increase in size the cyclopean image remains sharp, but binocular luster is also perceived.

Figure 38 Binocular combination demo. In this demonstration we reduced contrast in the monocular images by blurring local regions in the left and right eyes. In the right eye's image (on the left) the area just below the "A" in the corner of the book is blurred and in the left eye's image (on the right) the "L" in the banner is blurred. After fusing the monocular blur that is present in the monocular images becomes more difficult to detect in the cyclopean image.In the top row the blur is small and almost imperceptible, as the monocular blur regions increase in size the cylopean image becomes more distorted.

Guided by the ideas that contrast plays an important role in determining cyclopean appearance and that cyclopean appearance occurs at a local scale we devised three main experiments. In the first experiment we examined the role of global contrast on cyclopean appearance. Ding and Sperling (2006) have proposed a model that relies on the relative contrasts of the monocular images to determine how the visual system weights each eye's contribution to the cyclopean image; their model has been used to successfully predict cyclopean contrast for a wide variety of inter-ocular contrast differences (Huang et al. 2010, Hou et al. 2013). In experiment 1 we measured the perceived contrast of a cyclopean random dot pattern presented binocularly with different contrasts in each eye. Our second and third experiments were designed to study binocular combination at local scales. We used random white noise to present differing contrasts (experimental conditions included

both global differences between eyes and local differences within images) and presented these images to the two eyes; because the images have zero disparity and the underlying noise is the same in both eyes the images can be fused and binocular rivalry is eliminated.

General Methods

Apparatus. All experiments were conducted on a Macbook Pro 13" Retina Display using a single monitor mirror haploscope. The display resolution was 2560x1600 with a 60Hz refresh rate. The viewing distance was set to 128cm such that each pixel in the display subtended 0.3 minutes of arc. The maximum luminance of the display was 266 cd/m^2 and the minimum luminance was 0.4 cd/m^2. Gamma correction was applied to ensure that that luminance output was linear with pixel values. The experiment was run in a dark room so that the monitor provided the only measureable light input to the eyes. Vergence distance was set to match the viewing distance. Stimuli were rendered with PsychToolbox (Brainard 1997, Pelli 1997) in MATLAB.

Observers. Four subjects (one male, three female) 20-26 years old with normal or corrected-to-normal vision participated in all of the experiments. One was an author and the other three were naïve observers.

Contrast Modulated Stimuli. Contrast was defined using Michelson contrast, thus a 0% contrast grating is equivalent to a uniform grey field set to 50% of the monitor's maximum luminance. Using Michelson contrast ensured that the average luminance of each row (for horizontal contrast modulation) or column (for vertical contrast modulation) was the same regardless of contrast of the row or column. In all of the experiments contrast was presented using a white noise carrier. In the white noise image half of the pixels were designated as bright pixels and the other half as dark pixels. Contrast was generated by increasing and decreasing the luminance of the bright and dark pixels relative to 50% luminance. For example 20% contrast modulation was created by setting the bright pixels to 60% of the maximum luminance and the dark pixels to 40% of the maximum luminance.

In experiments with contrast modulation we varied contrast using a sinusoidal or square wave along either the azimuth or elevation of the carrier; the phase of the modulations was always random across trials. The randomized phase was either the same in both eyes (in-phase) or out-of-phase in the two eyes by 180°; the in-phase monocular images are identical in the two eyes. The underlying random noise carrier is identical in the two eyes for the out-of-phase modulations, but the overlaid contrast modulation is shifted by 180°. These contrast-modulated stimuli are advantageous because they allow us to study binocular fusion without the complications of binocular rivalry introduced by dissimilar stimuli (i.e. matching a dark pixel to a white pixel). The dissimilarities introduced by modulating image contrast out-of-phase to each other do not affect the underlying spatial structure in the retinal images, thus dark and bright pixels in one eye are always matched to a corresponding dark and bright pixels in the other eye at a different luminance.

Experiment 1: Perceived Cyclopean Contrast with Global Contrast Differences

We first measured perceived contrast of when the same noise carrier is presented binocularly with differing contrasts. This is similar to a situation where an individual has developed a cataract and has a low contrast image in the affected eye and a high contrast

image in the unaffected eye. We used a two-interval forced-choice paradigm to present a contrast-consistent ($C_{consistent}$) stimulus where the two eyes were presented random noise with matched contrast and a contrast-inconsistent stimulus (Figure 39) where the left and right eyes were presented with random noise of different contrast (C_L, C_R). A total of 10 different contrast-inconsistent pairings were examined, because we were worried about possible effects of eye-dominance we investigated 20 contrast-inconsistent combinations total (Table 1).

Left Eye Contrast (%)	Right Eye Contrast (%)
10	30, 50, 70, 90
30	10, 50, 70, 90
50	10, 30, 70, 90
70	10, 30, 50, 90
90	10, 30, 50, 70

Table 1 Contrast combinations examined in experiment 1. Twenty total combinations of C_L and C_R were examined; we measured 10 unique contrast conflict pairs, but 20 total conditions to account for the possibility of eye dominance (e.g. we measured perceived contrast for both C_R = 70%, C_L = 30% and C_R = 30%, C_L = 70%).

For each of the 20 contrast inconsistent conditions we varied the value of $C_{consistent}$ using the method of constant stimuli. We fit the resulting psychometric data with a cumulative Gaussian using a maximum-likelihood criterion (Fründ et al. 2011). The mean of that fitted function was the estimate of the value of $C_{consistent}$ for which observers reported $C_{consistent}$ appeared to have more contrast on 50% of the trials. The 20 contrast inconsistent conditions and were interleaved within a session and the presentation order of the contrast consistent and contrast inconsistent stimuli was also randomized.

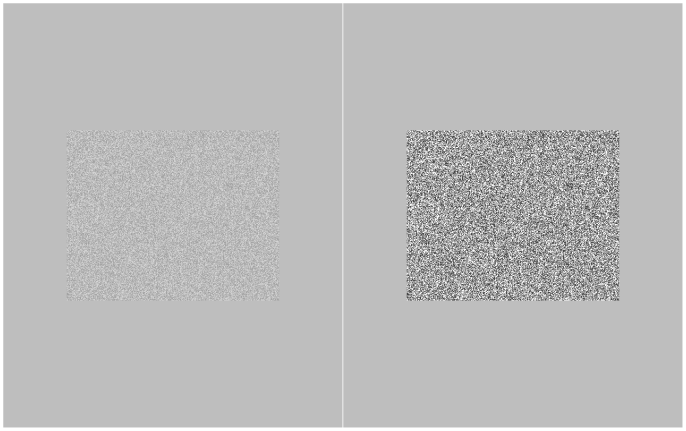


Figure 39 Contrast Inconsistent Stimulus. Stimuli were drawn on a uniform grey background set to 50% of the monitor's maximum luminance. All 20 combinations of C_R and C_L can be found in Table 1, the perceived contrast of each contrast inconsistent stimulus was determined through comparison to contrast consistent images using the method of constant stimuli.

Each of the contrast images was 3.2°x4.0° and each interval was two seconds long; prior to each stimulus presentation a gray subjects saw a gray screen set at 50% of the

monitor luminance with a dicoptic nonius cross. Upon stimulus onset the nonius cross disappeared and the contrast of the central 3.2°x4.0° in each eye increased from 0% to a specified value over 0.5 seconds. The contrast was then presented for one second and transitioned back to 0% over another 0.5 seconds. A one second wait interval with the fixation cross followed the first interval and the second interval presented following the same procedure. The random noise fields were 640 x 800 pixels, and as before there were an equal number of black and white pixels to ensure that the mean luminance of our stimuli were equal across all contrast values.

Data Analysis. Data was obtained using the method of constant stimuli and psychometric functions were fit to the data using psignifit (Fründ 2011). When the matched contrast of the monocular images was low, subjects would always report seeing the mismatched monocular images as having higher contrast (i.e. 0% of trials with the consistent contrast are perceived as having higher contrast). When the contrast of the matched images was high, subjects always report seeing the matched images as having higher contrast (i.e. 100% of trials with the consistent contrast are perceived as having higher contrast). We obtained the PSE of each condition outlined in Table 1 by fitting a psychometric function between 0% and 100% and determining the 50% response.

Experiment 1: Higher Contrast is More Heavily Weighted in Binocular Combination

First we consider three rules that might predict cyclopean contrast for our stimulus: using the maximum of C_L and C_R, using the average C_L and C_R, or using Ding and Sperling's (2006) gain-control of gain-control model. Panels B, C, and D from Figure 30 show these three predictions. The five missing points in the center of the plot are contrast consistent conditions where contrast was equal in both eyes; points that are farther away from the diagonal represent conditions where the contrast differences in the two eyes was most extreme. The perceived contrast using the maximum rule is a series of steps where the cyclopean contrast is perceived at 30%, 50%, 70%, and 90% depending on the combination of C_L and C_R. When using the average rule, for a given contrast value where C_L is larger than C_R, the maximum contrast is along the line where $C_R = C_L$ with decreasing contrast as C_R decreases relative to C_L. The opposite is true when $C_R > C_L$. Ding and Sperling's gain-control is similar to the average contrast rule when C_L and C_R are similar. However when there is a large difference between C_L and C_R the perceived contrast is biased heavily towards the larger of the two; interestingly, the predicted perceived contrast is non-monotonic at extreme contrast differences. For example, the predicted perceived contrast when C_L = 90% and C_R = 10% is larger compared to C_L =90% and C_R =30%.

Measured cyclopean contrast for the 20 contrast-conflict conditions are shown in Figure 30A and Figure 41 (n = 4, 9890 total trials). The root mean square error for the Ding and Sperling model was 3.4% while the maximum and average contrast rules yielded RMSEs of 12.4% and 11.6% respectively. In 20/20 contrast combinations the perceived cyclopean contrast was greater than the average contrast of the two eyes; in 17/20 conditions this difference was statistically significant (Figure 41). Of these three conditions, two were conflicts where the difference between the average contrast and gain-control predictions were smallest (C_L =70%, C_R = 90% and C_R =70%, C_L = 90%); the third was C_L =70%, C_R = 30%. In every condition the perceived contrast is greater than the average (red

points in Figure 41), i.e. biased towards the higher contrast of C_L and C_R. This bias is clearly demonstrated at large inter-ocular contrast differences when the average contrast is a very poor predictor of the perceived contrast compared to the Ding and Sperling model (black points in Figure 41). We found that 13/20 conditions agreed with their model. The gain-control model overestimated perceived contrast in five conditions (C_L/C_R: 90/10, 50/10, 70/10, 70/30, 90/10) and underestimated perceived contrast in two (30/70 and 50/90).

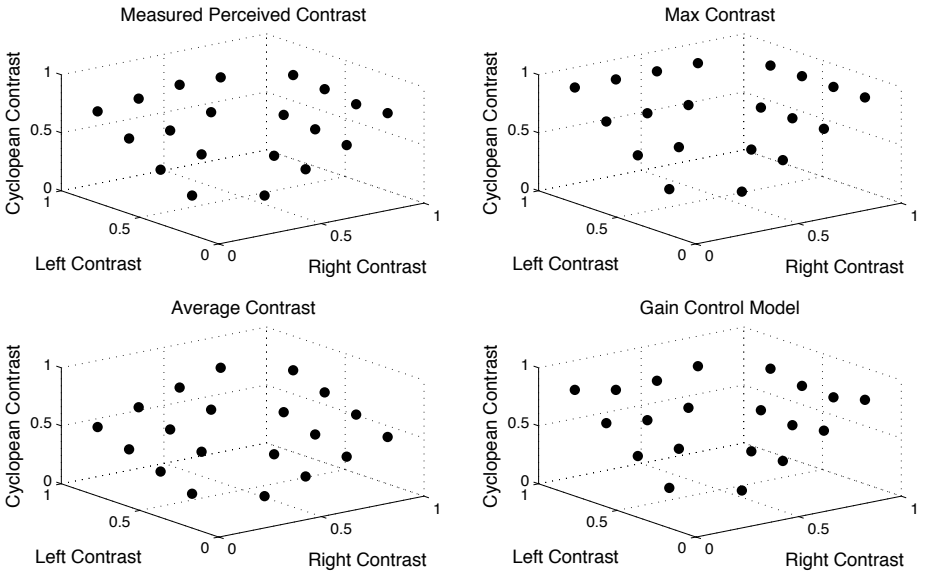


Figure 40 Experiment 1 Results. (A) The perceived cyclopean contrast (z-axis) is plotted as a function of the contrasts in the left (y-axis) and right (x-axis) eyes. Panels B, C, and D represent predictions of cyclopean contrast if observers utilize the maximum contrast in the two eyes (B), the average contrast presented to both eyes (C), and the gain-control of gain-control model proposed by Ding and Sperling.

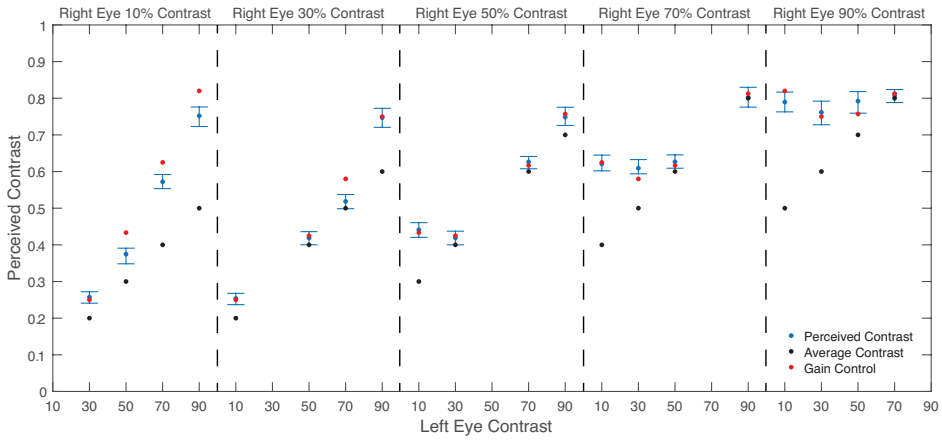


Figure 41 Perceived Cyclopean Contrast. Blue dots and error bars (95% CI) represent subject's perceived cyclopean contrast. Red dots represent the average contrast presented to the left and right eyes and black dots are predictions generated using Ding and Sperling's gain-control model.

There does not appear to a strong eye-dominance effect. When we compared the 10 contrast conflict conditions with C_L and C_R reversed the differences in cyclopean contrast were small, ranging from 0.4% to 8.7% with an average difference of 3.7%. The largest differences occur when the contrast of one eye is high and the contrast of the other eye is low. Of the 10 conditions reversing the contrasts for there conditions resulted in a significant difference (C_L/C_R = 10/50, 10/70, and 30/70).

Left Eye Contrast (%)	Right Eye Contrast (%)	Difference after Contrast Reversal (%)
10	30	1.0
10	50	**8.5**
30	50	1.1
10	70	**7.6**
30	70	**8.7**
50	70	0.6
10	90	4.1
30	90	1.8
50	90	3.4
70	90	0.4

Table 2 Effect of eye-dominance on perceived contrast. The right column is the difference between the perceived contrast when the contrasts of the left and right eyes are switched. For example, in the first row perceived cyclopean contrast when the left eye's image is 10% contrast and the right eye's image is 30% appears to be 1% higher than the reverse condition when the left contrast is 30% and the right eye contrast is 10%. The three values highlighted in bold indicate conditions where there was a significant difference after switching the contrast presented to the left and right eyes.

Experiment 2: Role of Local Contrast differences in Cyclopean Perception

In the presence of inter-ocular contrast differences, the visual system gives weight to the retinal image with higher contrast. Ding and Sperling's non-linear gain-control model assigns different weights to the two eyes depending on their relative contrasts and the model predicts perceived cyclopean contrast fairly accurately. However, their model was designed using a global weighting that is applied to the entire retinal image of the left and right eyes , but evidence suggests that the cyclopean image can be formed as a local combination of the retinal images. With this in mind, we set out to determine if the cyclopean image resulting from binocular fusion could be conclusively shown to occur at a local, rather than global, scale.

To address this question the question of spatial integration we compared contrast modulation detection threshold for in-phase and 180° out-of-phase binocular contrast modulation gratings of different spatial frequencies. A 1000x1000 pixel (5°x5°) white noise image was used as a carrier to present contrast variation (see general methods for more detail); contrast modulations were sinusoidal and were presented vertically or horizontally. Prior to and after each trial subjects were shown a gray background

equivalent to 50% of the screen's maximum luminance, the background also included a zero-disparity border as a vergence lock (similar to Figure 45) with a dicoptic nonius. Subjects were instructed to align the vertical halves of the nonius and initiate a trial via key press after reaching proper vergence (set to the screen distance). During each trial a random noise pattern with vertical or horizontal contrast modulation would, after one second the random noise field was replaced with the background image and subjects were asked to report whether the perceived contrast modulation was vertical or horizontal.

In-phase and out-of-phase modulations at spatial frequencies ranging from 0.19-5.9cpd were interleaved across trials; the phase of the modulation relative to the top of the image frame was random in every trial. The maximum contrast was always set to 90% and increasing contrast modulation amplitude decreased the minimum contrast (i.e. a 40% contrast modulation resulted in a minimum contrast of 50%).

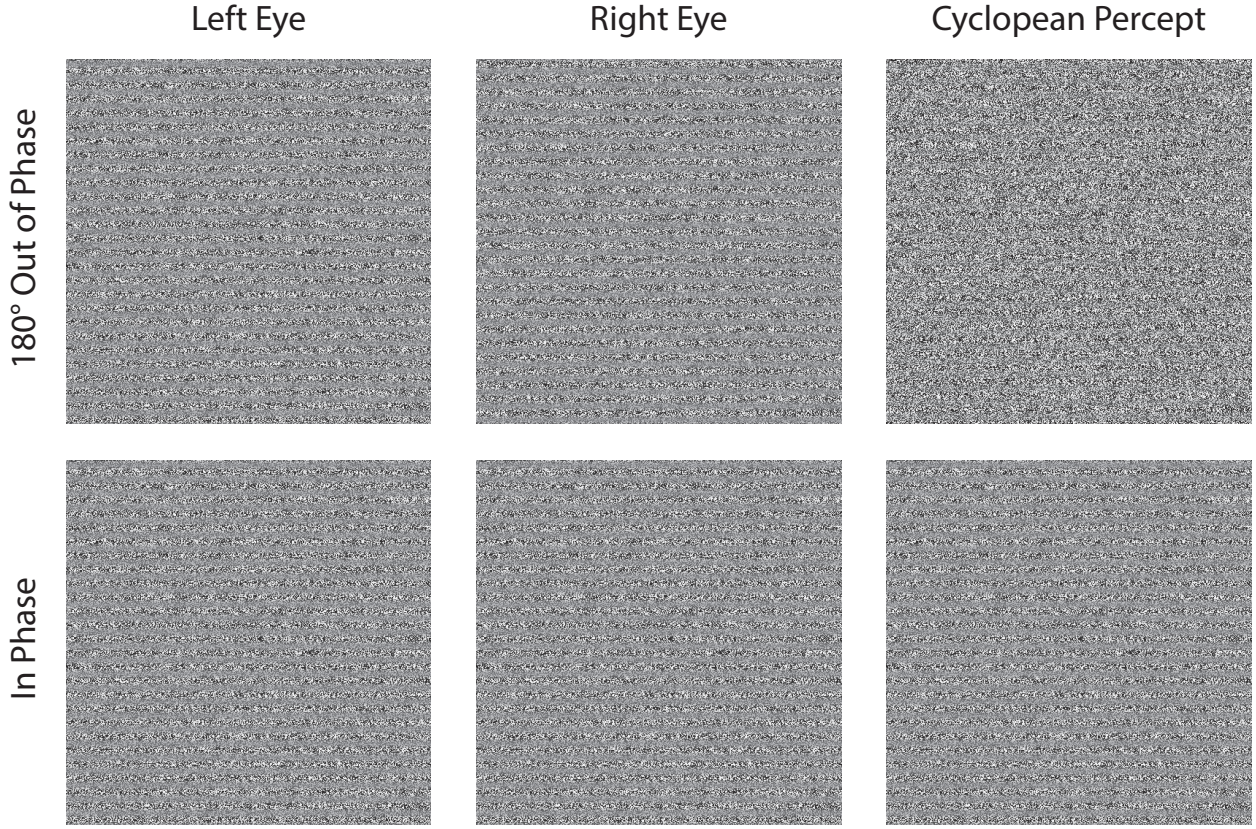


Figure 42 Cyclopean Contrast Modulation Stimulus. The top and bottom rows represent a stimulus that is in-phase and out-of-phase respectively. Each column represents the left eye's image, the right eye's image, and the simulated cyclopean percept after binocular combination. A strong vergence lock was drawn around each monocular image, but is not shown in this figure. A larger version of the border can be seen in **Figure 45**.

Data Analysis. When contrast modulation amplitudes were low subjects perceived random noise of uniform contrast and correctly identified contrast modulation orientation at chance (50%). Performance increased monotonically with contrast modulation amplitude until subjects were correct on 100% of trials. To determine contrast modulation detection threshold we fit a psychometric curve from 50% to 100% and found the 75% correct mark. We measured the contrast modulation detectability thresholds for spatial

frequencies from 0.19-5.9cpd and compared the difference in contrast modulation threshold between in-phase and out-of-phase gratings to determine whether or not the visual system would create a cyclopean percept which maximized perceived contrast across the monocular images.

Short Duration Vergence Experiment. We wanted to investigate the potential effects of vergence eye movements on our results. So we presented the stimulus for 150ms a duration too short for vergence eye-movements and reduced the size of our stimulus to 4°x4° to make the vergence border stronger. We were also interested in investigating potential differences between detection of horizontal and vertical modulations, so instead of utilizing a 2AFC paradigm we changed the experiment to a yes/no detection task and calculated d-prime for vertical and horizontal gratings that were either in-phase or out-of-phase for 0.5 and 5.9cpd gratings at contrast modulations from 0-70%.

Experiment 2: Binocular Combination is a Local Phenomenon

In order to detect the presence of a sinusoidal contrast modulation the visual system must be able to perceive a difference between the highest and lowest contrast regions of the modulation. The in-phase monocular gratings are dioptic and they will be combined with no change in the cyclopean image, thus the contrast modulation detection magnitude is simply the contrast modulation detection threshold. If cyclopean contrast is derived locally from monocular contrasts of the two eyes, then the contrast differences between the left- and right-eyes for 180° out-of-phase contrast modulation will be combined in the cyclopean image that will be derived from both retinal contrasts and should be lower for 180° out-of-phase modulations compared to in-phase modulations. In other words, for a given contrast modulation amplitude, the apparent contrast in the out-of-phase conditions will appear more uniform compared to the same contrast modulation in the in-phase conditions and more difficult to detect. To measure this effect we presented a vertical or horizontal contrast modulation and asked subjects to identify the orientation of the modulation; when subjects cannot detect contrast modulation, the cyclopean image is uniform contrast and subjects correctly report the orientation of the modulation at chance. We then compared the contrast modulation thresholds for in-phase and out-of-phase gratings to see if detection thresholds were higher for out-of-phase modulations.

In addition to a threshold difference between in-phase and out-of-phase gratings, local binocular combination might also fail at a relatively low spatial resolution. This is because such binocular a mechanism would require a binocular neuron to compare regions from the left and right eyes (similar to stereopsis). Such a neuron would have a relatively large receptive field and the size of this field would place limitations on the finest stimulus that could be fused across both eyes in a manner consistent with the mechanisms observed in stereopsis. To study this effect we compared contrast detection thresholds over various spatial frequencies to determine what the size of such a binocular receptive field might be.

The results of our experiment are presented in Figure 43 (n=4, 23436 total trials), in-phase detection thresholds are shown in red and out-of-phase thresholds are shown in blue. Below approximately 6cpd the detection thresholds for out-of-phase gratings higher than in-phase gratings demonstrating that the apparent contrast for out-of-phase gratings does appear more uniform. This difference must be attributed to binocular combination because the underlying monocular images in the in-phase and out-of-phase conditions

possess the same image statistics and the only difference is their relative phase. As spatial frequency increases, the detection thresholds begin to converge and meet at approximately 6cpd suggesting that a binocular mechanism exists which integrates left- and right-eye monocular images over a spatial region of approximately 10 minutes of arc.

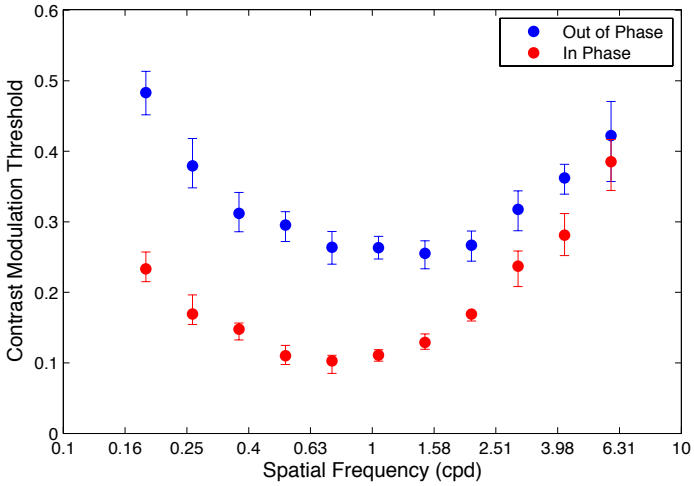


Figure 43 Contrast Modulation Threshold of Contrast Modulated Gratings. The abscissa represents spatial frequency of the sinusoidal contrast modulation and the ordinate represents the contrast modulation detection threshold. Up to approximately 6cpd contrast modulation detection threshold is higher for out-of-phase gratings suggesting demonstrating the effect of binocular fusion on the cyclopean appearance of the contrast modulated gratings.

Verification of Spatial Frequency Limit. We postulate that the convergence of in-phase and out-of-phase contrast modulation thresholds in Figure 43 can be attributed a binocular mechanism which combines local regions of the retinal images. Alternatively, the visual system is making vergence eye movements to align high contrast regions present in both eyes negating any binocular differences in the in- and out-of-phase conditions. To investigate this possibility we repeated the main experiment with 150ms presentation time (too short for vergence eye movements) and 20% smaller stimulus to increase the strength of the vergence lock border. We also ran the experiment as a yes/no detection task and measured d-prime to examine potential differences arising from differences in vergence eye movements for horizontal and vertical contrast modulations.

The results of the short duration experiment are shown below in Figure 44. The top row represents contrast modulations at 0.5cpd and the bottom row represents data collected at the spatial frequency limit where the thresholds converged at 5.9cpd. Thresholds extracted from d-prime are usually determined when d-prime is equal to one and thresholds calculated using this metric roughly agree with the thresholds from experiment 2 (0.5cpd 11% in-phase/30% out-of-phase, 5.9cpd 39% in-phase/42% out-of-phase). Although horizontal contrast modulation appears to be easier to detect for the in-phase gratings, there is no difference between vertical and horizontal orientations when the contrast modulation is out-of-phase. We can rule out the effects of vergence eye movements as the cause for convergence of contrast modulation detection thresholds at 5.9cpd because the thresholds in this follow-up study are in agreement with the data

obtained from the main experiment and there appears to be no difference in orientation for the out-of-phase gratings

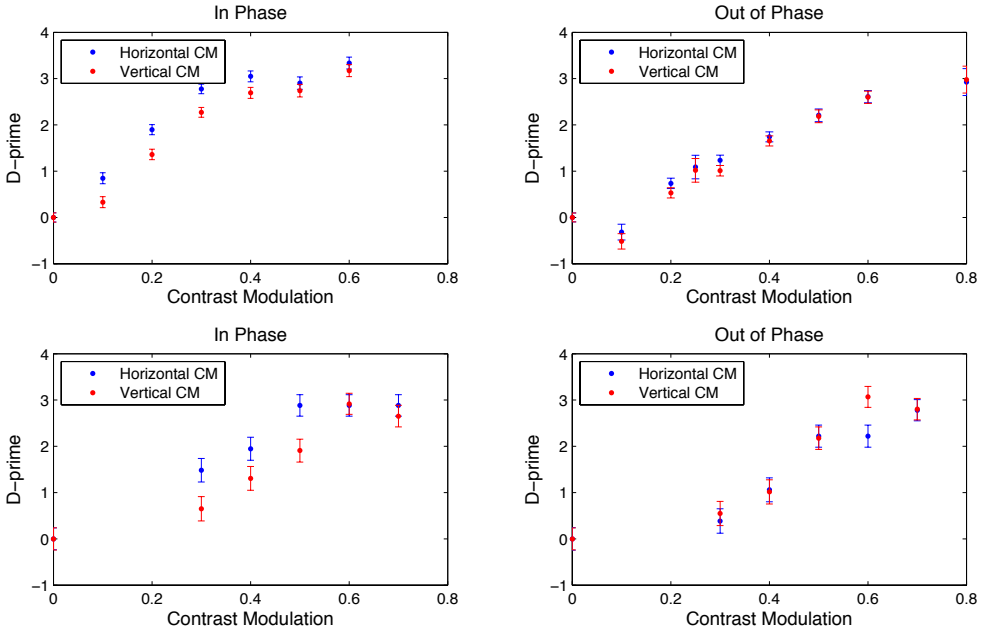


Figure 44 D-prime after eliminating vergence eye movements. Top row is 0.5cpd, bottom row is 5.9cpd. Left column is in-phase, right column is out-of-phase. Threshold obtained from d-prime equal to one is similar to results from previous experiment. No difference in vertical and horizontal modulation for out-of-phase.

Experiment 3: Perceived Cyclopean Contrast with Local Contrast Differences

In the previous experiment we measured the apparent contrast of contrast-inconsistent binocular images. The out-of-phase contrast modulated stimuli used in experiment 2 are also contrast-inconsistent and we wanted to compare the perceived contrast of out-of-phase stimuli to the contrast-inconsistent stimuli used above. We used square wave contrast modulation rather than sine wave modulation in order to compare these results with perceived contrast values obtained in experiment 1.

In this experiment we utilized a two-alternative forced-choice paradigm rather than a two-interval forced-choice. Prior and after each trial subjects were presented a 50% luminance gray background as well as a zero-disparity vergence lock and dichoptic nonius cross. For each trial we presented a two 3°x3° (600x600 pixel) contrast patches, one without contrast modulation and the other with an out-of-phase square wave modulation; trial duration was one second. One image was placed above the nonius cross and the other below, an equal number randomly interleaved trials were run with the contrast modulation above and below the nonius cross. Subjects were asked to identify whether the image above or below the fixation cross had higher apparent contrast. The contrast modulations were presented at either 0.5 or 2.09cpd with a maximum contrast of 75% or 90% and a contrast modulation of either 20% of 40% resulting in a total of eight different conditions.

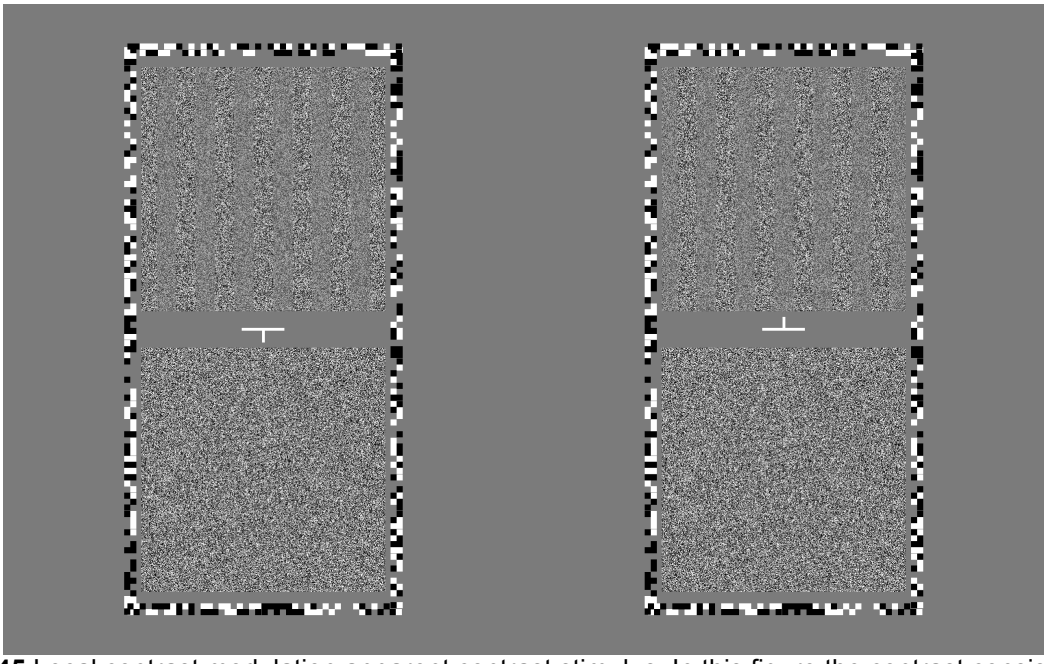


Figure 45 Local contrast modulation apparent contrast stimulus. In this figure the contrast consistent stimulus is on the bottom and the contrast-modulated stimulus is on top. For a given contrast modulated stimulus we used the method of constant stimuli to vary the contrast consistent stimulus to find the point of subjective equality.

Data Analysis. Data was obtained using the method of constant stimuli and psychometric functions were fit to the data using psignifit. When the contrast the uniform fields was low, subjects would always report seeing the contrast modulated cyclopean image as having higher contrast (i.e. 0% of trials with uniform contrast are perceived as having higher contrast). When the contrast of the uniform images was high, subjects always report seeing the uniform images as having higher contrast (i.e. 100% of trials with the uniform contrast are perceived as having higher contrast). We obtained the PSE of each condition eight conditions by fitting a psychometric function between 0% and 100% and determining the 50% response.

Experiment 3: Local Contrast Calculation Biased Towards Higher Contrast

In experiment 1 Ding and Sperling's gain-control model was able to account for global gain of the monocular images to predict perceived contrast. The findings from experiment 2 show that cyclopean contrast appears more uniform in the out-of-phase case, however, the results do not allow for conclusions about how the left- and right-eye images are combined. It is possible that a global weighting like the one used in Ding and Sperling's gain-control model could account for the local contrast combination observed in experiment 2. If weighting in the gain-control model is determined by measuring the overall contrast of a retinal image, then both retinal images would receive equal weighting and the cyclopean contrast at every point would be the average contrast of the two retinal images. Averaging would halve the apparent cyclopean contrast relative to the maximum contrast in each monocular image. Such a weighting would also imply that the visual system uses a global comparison of the retinal image contrasts rather than local weighting that varies according to inter-ocular contrast differences. Alternatively the visual system

might weight image regions locally according to the Ding and Sperling model and in this case the cyclopean contrast would be near the maximum contrast in the contrast modulation. To investigate these two distinct possibilities we measured the perceived contrast of 180° out-of-phase contrast modulated square wave gratings. We elected to use square wave modulation over sinusoidal modulation so that the results from this experiment could be more readily compared to the results of experiment 1.

Apparent contrast for 0.5cpd and 5.9cpd modulations with maximum contrasts of 75% and 90% are show in Figure 46 (n=4, 6524 trials); we examined the effect of vertical and horizontal orientations and, like the short-duration vergence experiment, found no effect of orientation. The following data are the combined trials of both horizontal and vertical orientations. The left plot shows apparent contrast for square wave gratings with a maximum contrast of 75% and the right plot shows apparent contrast for gratings with 90% maximum contrast. For each maximum contrast two spatial frequencies (0.5cpd and 5.9cpd) were observed along with two contrast modulations (20% and 40%). The minimum contrast and spatial frequencies are shown in the abscissa and apparent contrast is plotted along the ordinate. Observer data with 95% confidence intervals are show in blue, predictions using the average contrast are shown in black, and predictions using the gain-control model are shown in red. Like orientation, spatial frequency does not appear to have an effect on apparent contrast. For both 75% and 90% maximum contrast the average contrast and gain-control predictions are similar for the smallest contrast modulation of 20%. Despite these similarities, in three out of four conditions at 20% contrast modulation, the average contrast prediction appears to be a worse predictor than the gain-control model. A much larger difference exists in the predictions at 40% contrast modulation and the results here are more revealing. In these conditions the observed contrast is much higher than the average contrast and even the gain-control model appears to under predict the perceived contrast. Like experiment 1 the Ding and Sperling gain-control model is able to predict the perceived contrast and furthermore because these waveforms have locally varying contrast, the gain-control model must occur at a local, not global, scale. In conjunction with the results from experiment 2 these data show that contrast is generated at local scale between both eyes and that the relative weights of the left and right eyes changes depending on local image contrast.

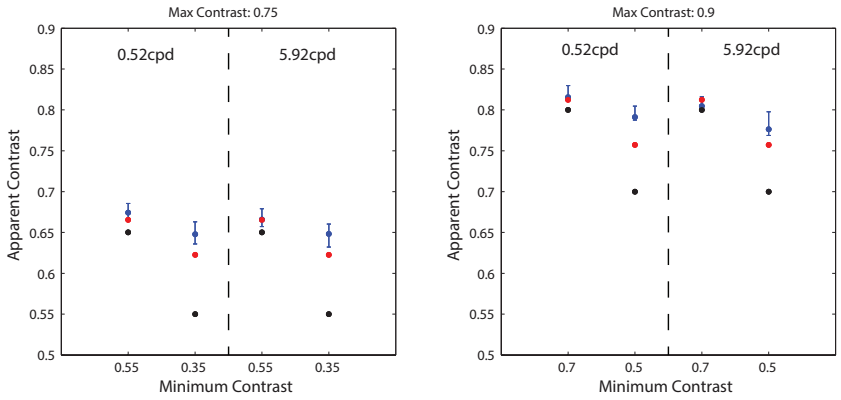


Figure 46 Apparent contrast of 180° out-of-phase square wave gratings, there was no different in grating orientation so trials for both vertical and horizontal orientations are combined. For small contrast modulations (20%) the difference between the average contrast and gain-control predictions are small so we also measured a larger contrast modulation of 40%.

Discussion

Existing Models of Binocular Combination

Ding and Sperling (2006) used perceived phase from two monocular phase-offset luminance gratings to measure the relative contribution of each eye. Huang et al. (2010) used a similar experiment this time measuring both perceived contrast and perceived phase. They found that perceived contrast and phase of the cyclopean images are independent and proposed a model where the weighting of each monocular image is independent to the perceived phase. By using sinusoidal luminance gratings both of these studies were restricted to measuring the relative weight of each eye globally, because our stimuli used a random noise carrier, we were able to define contrast at a local scale and investigate binocular combination weighting at a local scale. Zhou et al. (2014) used a similar contrast grating to investigate the perceived phase of gratings with different phases.

By examining the perceived cyclopean image when gratings were 180 degree out of phase, we demonstrated in experiment 2 that information from the monocular images is lost in the cyclopean image. Subjects are not able to discriminate the orientation of the contrast grating even when the monocular contrast modulation is above detection threshold. The use of contrast gratings was essential in this experiment because using luminance gratings would have resulted in binocular rivalry, contrast gratings do not induce rivalry because there the random dot carrier has corresponding points in both eyes. By measuring the effect as a function of spatial frequency we were also able to determine the scale at which this effect occurs.

Relationship to Stereopsis and Local-Cross Correlation

Disparity detection is dependent upon the contrast and spatial frequency content of the underlying monocular images (Frisby and Mayhew 1978, Schor and Wood 1983, Halpern and Blake 1988, Legge and Gu 1989). In general, reducing the contrast of the monocular images reduces disparity sensitivity. Interestingly, if only one eye's image is reduced in contrast, stereoacuity is worse than if both eyes' contrasts were reduced by the same amount; this is known as the contrast paradox (Simons 1984, Schor and Heckman 1986, Stevenson and Cormack 2000, Ding and Levi 2011).

Several models have examined the effects of monocular contrast on local-cross correlation (Legge and Gu 1989, Kontsevich and Tyler 1994, Cormack et al. 1997). Hou et al. (2013) implemented a modified version of Ding and Sperling's contrast-gain model and were able to predict increases in disparity thresholds based on differences in weighting. In experiments 2 and 3 our results indicate that the visual system weights monocular images by their local inter-ocular contrast differences to maximize the perceived cyclopean contrast. The physiological basis for this behavior requires binocular neurons with relatively large receptive field because the spatial frequency cutoff of the effect is low, just 6cpd. It would be interesting to create random dot stereograms using our stimuli to see what effects local, rather than global, differences in contrast have on stereopsis in order to gain insight into the mechanisms behind both phenomena. For example, is it possible that local regions with large inter-ocular contrast differences have degraded depth from disparity, but can be "filled-in" by the larger surrounding regions that have more similar contrasts; what percentage and distribution of matched and mismatched inter-ocular contrasts might lead to increased disparity thresholds?

Applications in Head-Mounted and Stereo3D Displays

One potential application of stereo3D displays is to use low dynamic range stereo images to create the appearance of a high dynamic range (HDR) image (Sun et al. 2010, Xuan et al. 2012). This method relies on the visual system to combine the stereoimages so that the cyclopean image appears to have more dynamic range than either monocular image. Our results from experiment 2 demonstrate that the visual system does merge higher contrast regions in the left- and right-eye monocular images to create a cyclopean image. However, the threshold criteria in our discrimination task is quite stringent because any detectable non-uniformity, even if a majority of the cyclopean image appears uniform, allows subjects to judge orientation. This means that the tolerance for inter-ocular contrast difference is potentially higher than the 30%-50% modulation outlined in Figure 43. In Figure 46, for example, the apparent contrast of the cyclopean image remains high even when the monocular contrast modulation amplitude is above threshold for 0.5cpd and near threshold for 5.9cpd at 40%. This means that fusion artifacts are visible at the transition zones between the left- and right-eye images, but the average contrast remains high. Insight into how visible artifacts at transitions between the monocular images might be minimized is related spatial integration. Our findings from Figure 43 show that the left- and right-eye are able to combine gratings up to approximately 6cpd. If we translate this to a rough window size that is a value of 10 minutes of arc, i.e. the transition between desired regions in one eye to the other should take place over a scale of approximately 10 minutes of arc in order to reduce fusion artifacts in the cyclopean image.

Conclusions

The question of how the visual system combines retinal images into a single, coherent percept of the world is critical to understanding visual perception. Previous work in rivalry has attempted to study the mechanisms that underlie this process by presenting conflicting images to both eyes; these rivalrous stimuli are unrepresentative of local differences in the retinal images that would occur from viewing real scenes. We studied how local ambiguities within left and right retinal images might be resolved based on

image contrast and present three main findings. First, we found that the Ding and Sperling gain-control model predicted apparent contrast for non-rivalrous stimuli with global inter-ocular contrast differences. Second, we determined that the cyclopean image is a constructed from local comparisons between the left and right eyes and not a global weighting of the two eyes. Finally we determined that the local combination of left and right eyes is weighted according to image contrast measured at a local, not global scale. Together these results show that the visual system is able to combine monocular images into a cyclopean percept that preserves contrast throughout the entire image and provide insight into how the visual system is able to resolve ambiguities between retinal images that commonly arise in natural viewing from viewing reflections and occlusion boundaries.

References

Arnold D (2011). Why is binocular rivalry uncommon? Discrepant monocular images in the real world. *Frontiers in Human Neuroscience*, 5, 116, 1-7.

Arnold D, Grove P, Wallis T (2007). Staying focused: A functional account of perceptual suppression during binocular rivalry. *Journal of Vision*, 7(7), 7.1–8

Brainard, DH (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.

Banks MS, Gepshtein S, and Landy MS. (2004), What is Spatial Stereoresolution So Low? *J. Neuro.*, 24 (9), 2077-2089.

Brascamp JW, Klink PC, Levelt WJ (2015). The 'laws' of binocular rivalry: 50 years of Levelt's propositions. *Vision Res.,* 109(Pt A):20-37.

Cormack LK, Stevenson SB, Landers DD (1997). Interactions of spatial frequency and unequal monocular contrasts in stereopsis. *Perception,* 26(9), 1121-1136.

Cormack LK, Stevenson SB, Schor CM. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Res*., 31(12), 2195-207.

Ding J and Levi D (2011). Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision. *PNAS,* 108(37) E733-741.

Ding, J and Sperling, G. (2006). A gain-control theory of binocular combination. *PNAS*, 103 (4), 1141-1146.

Fahle M (1982). Binocular rivalry: suppression depends on orientation and spatial frequency. *Vision Res.*, 22(7):787-800.

Fleet DJ, Wagner H, Heeger DJ. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res.*, 36(12), 1839-57.

Filippini HR and Banks MS. (2009). Limits of stereopsis explained by local cross-correlation. *J. Vis.*, 9 (1), 8.1-818.

Mayhew JE, Frisby JP (1978). Contrast summation effects and stereopsis. *Perception,* 7(5):537-50.

Fründ, I, Haenel, NV, Wichmann, FA. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *J. Vision*, 11 (16).

Halpern, DL and Blake, RR. (1988). How contrast affects stereoacuity. *Perception*, 17, 483–495.

Harris JM, McKee SP, Smallman HS. (1997). Fine-scale processing in human binocular stereopsis. *J Opt Soc Am A Opt Image Sci Vis*., 14(8), 1673-83.

Hou F, Huang CB, Liang J, Zhou Y, Lu ZL (2013). Contrast gain-control in stereo depth and cyclopean contrast perception. *J. Vision,* 13(8):3, 1-19.

Huang CB, Zhou J, Zhou Y, Lu ZL. (2010) Contrast and Phase Combination in Binocular Vision. *PLoS ONE*, 5(12), e15075.

Kane D, Guan P, Banks MS. (2014). The Limits of Human Stereopsis in Space and Time. *J. Neuro.*, 34 (4), 1397-1408.

Kontsevic LL and Tyler CW (1994). Analysis of stereothresholds for stimuli below 2.5c/deg. *Vision Res.,* 34(17), 2317-2329.

Legge GE and Gu YC. (1989). Stereopsis and contrast. *Vis. Res.*, 29 (8), 989–1004.

O'Shea R (2011). Binocular rivalry stimuli are common but rivalry is not. *Frontiers in Human Neuroscience*, 5 (2011), pp. 1–2

Pelli, D.G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10:437-442

Read JCA, Parker AJ, and Cumming BG (2002). A simple model accounts for the reduced response of disparity-tuned V1 neurons to anti-correlated images. *Vis. Neurosci.,* 19, 735-753.

Stevenson, S and Cormack, K. (2000). A contrast paradox in stereopsis, motion detection, and Vernier acuity. *Vision Res.,* 40 (21), 2881-4.

Schor C, Heckmann T (1989). Interocular differences in contrast and spatial frequency: effects on stereopsis and fusion. *Vision Res.,* 29(7):837-47.

Schor MC and Wood I (1983). Disparity Range for Local Stereopsis as a Function of Luminance Spatial Frequency. *Vision Res,* 23(12) 1649-1654.

Simons K (1984). Effects on stereopsis of monocular versus binocular degradation of image contrast. *Inves. Opth. & Visual Science,* 25(8), 987-989.

Sun N, Mansour H, Ward R (2010). HDR image construction from multi-exposed stereo LDR images. *Image Processing (ICIP), 2010 17th IEEE International Conference on* pp.2973,2976.

Tong F, Ming M, Blake RR. (2006). Neural bases of binocular rivalry. *Trends in Cog. Sci*., 10 (11), 502-511.

Wolfe JM (1983). Influence of spatial frequency, luminance, and duration on binocular rivalry and abnormal fusion of briefly presented dichoptic stimuli. *Perception,* 12(4):447-56.

Wheatstone C (1838). Contributions to the physiology of vision. -Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London* 182:371-394.

Xuan, Y, Linling Z, Tien-Tsin W, and Pheng-Ann H. (2012). Binocular tone mapping. *ACM Trans. Graph.*, 31 (4), Article 93

Zhou J, Liu R, Zhou Y, Hess RF (2014) Binocular Combination of Second-Order Stimuli. *PLoS ONE*, 9(1): e84632.