

**Perceptual artifacts associated with novel display technologies**

by

Paul Vincent Johnson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Jt. Doctor of Philosophy

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

and

University of California, San Francisco

Committee in charge:

Professor Martin S. Banks, UC Berkeley, Chair

Professor Christoph Schreiner, UCSF

Associate Professor David Whitney, UC Berkeley

Spring 2015

**Perceptual artifacts associated with novel display technologies**

Copyright 2015  
by  
Paul Vincent Johnson

## Abstract

Perceptual artifacts associated with novel display technologies

by

Paul Vincent Johnson

Jt. Doctor of Philosophy in Bioengineering

University of California, Berkeley

and

University of California, San Francisco

Professor Martin S. Banks, UC Berkeley, Chair

Stereoscopic 3D displays are able to provide an added sense of depth compared to traditional displays by sending slightly different images to each eye. Although stereoscopic displays can provide a more immersive viewing experience, existing methods have drawbacks that can detract from image quality and cause perceptual artifacts. In this thesis I investigate perceptual artifacts associated with displays, and propose novel techniques that can improve viewing experience compared to existing methods. Chapter 1 presents a broad introduction to the various types of artifacts that can occur in displays, including motion artifacts, depth distortion, flicker, and color breakup. In Chapter 2, I describe a novel display technique, “spatiotemporal interlacing,” that combines spatial and temporal interlacing. I demonstrate using psychophysics that this method provides a better viewing experience than existing methods, and I present a computational model that confirms the psychophysical data. In Chapter 3, I present an investigation of perceptual artifacts on a high-frame-rate (240Hz) temporally interlaced OLED display. The high frame rate of this display allows for several unique driving modes. I investigated the perceptual consequences of these driving modes, characterizing the display in terms of motion, depth distortion, flicker, spatial resolution, and luminance. I demonstrate how one’s selection of viewing mode can tailor the viewing experience depending on the goals of viewer. Chapter 4 discusses the phenomenon of color breakup, a perceptual artifact that occurs in displays that present colors sequentially in time, such as Digital Light Processing (DLP) projectors. I discuss a novel psychometric procedure to measure the artifact, and a way to model the saliency of the artifact based on known spatiotemporal properties of the human visual system. I also propose a method for reducing color breakup.

# Contents

<b>Contents</b>	<b>i</b>
<b>List of Figures</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Existing stereoscopic 3D techniques . . . . .	1
1.2 Motion artifacts and flicker . . . . .	5
1.3 Depth distortion . . . . .	9
1.4 Spatial resolution . . . . .	11
1.5 Color breakup . . . . .	13
1.6 Goals . . . . .	13
<b>2 Spatiotemporal hybrid display</b>	<b>15</b>
2.1 Introduction . . . . .	15
2.2 Experiment 1: Motion artifacts . . . . .	18
2.3 Experiment 2: Flicker . . . . .	24
2.4 Experiment 3: Spatial resolution . . . . .	29
2.5 Experiment 4: Depth distortion . . . . .	33
2.6 Discussion . . . . .	34
2.7 Conclusion . . . . .	42
<b>3 Perceptual artifacts on a 240Hz OLED display</b>	<b>43</b>
3.1 Introduction . . . . .	43
3.2 Experiment 1: Motion artifacts . . . . .	44
3.3 Experiment 2: Flicker . . . . .	48
3.4 Experiment 3: Depth distortion . . . . .	50
3.5 Discussion . . . . .	53
3.6 Impact . . . . .	57
3.7 Conclusion . . . . .	58
<b>4 The visibility of color breakup</b>	<b>59</b>
4.1 Introduction . . . . .	59

4.2	Model of color breakup saliency . . . . .	62
4.3	Psychophysical methods . . . . .	68
4.4	Results . . . . .	71
4.5	Discussion . . . . .	75
<b>5</b>	<b>Conclusion</b>	<b>79</b>
	<b>Bibliography</b>	<b>81</b>

# List of Figures

1.1	Haploscopes used in vision science research . . . . .	2
1.2	Stereoscopic 3D display protocols . . . . .	3
1.3	Retinal-image stimulation with different display protocols . . . . .	6
1.4	Effect of stroboscopic sampling on amplitude spectrum . . . . .	8
1.5	Depth distortion in temporally interlaced displays . . . . .	10
1.6	Average visual acuity for different interlacing techniques. . . . .	12
1.7	Appearance of color breakup . . . . .	14
2.1	Temporal, spatial, and hybrid interlacing . . . . .	16
2.2	S3D display protocols schematized in space-time plots . . . . .	17
2.3	Calculated luminance modulation for CRT and sample-and-hold displays . . . . .	20
2.4	Stimulus used to measure visibility of motion artifacts . . . . .	20
2.5	Motion artifact thresholds as a function of frame rate . . . . .	21
2.6	Motion artifact thresholds as a function of capture rate . . . . .	23
2.7	Visibility of motion artifacts with tracking eye movements . . . . .	24
2.8	Motion artifacts and presentation time . . . . .	25
2.9	Flicker visibility for different protocols and block heights . . . . .	27
2.10	Stimulus used to measure spatial resolution . . . . .	30
2.11	Threshold viewing angle as a function of pixel size . . . . .	31
2.12	Effect of motion on spatial resolution . . . . .	32
2.13	Stimulus used to measure depth distortion . . . . .	34
2.14	Depth distortion results . . . . .	35
2.15	Simulation of hybrid technique . . . . .	38
2.16	Amplitude spectra for different interlacing protocols (simulation) . . . . .	39
2.17	Amplitude spectra for different interlacing protocols (cartoon) . . . . .	40
2.18	Implementing the hybrid technique . . . . .	41
3.1	Driving modes presented on the 240Hz display . . . . .	45
3.2	Method used to measure motion artifacts . . . . .	46
3.3	Motion artifact results, individual data . . . . .	47
3.4	Effect of flash number on motion artifacts . . . . .	48
3.5	Effect of duty cycle on motion artifacts . . . . .	49

3.6	Effect of interocular delay on motion artifacts . . . . .	50
3.7	Flicker thresholds averaged across the subjects . . . . .	51
3.8	Effect of interocular delay on depth distortion . . . . .	52
3.9	Effect of eye movements on the perception of judder and blur . . . . .	54
4.1	Space-time plot of retinal images . . . . .	60
4.2	Space-time differences in color modulation . . . . .	61
4.3	Space-time color modulation with offset . . . . .	63
4.4	Model of color breakup . . . . .	64
4.5	Model predictions . . . . .	67
4.6	Protocols tested in color-breakup experiment . . . . .	68
4.7	Stimulus used to measure color breakup . . . . .	69
4.8	Psychometric function for perception of color breakup . . . . .	70
4.9	Color breakup when subjects tracked a yellow or magenta stimulus . . . . .	72
4.10	Color breakup, individual data . . . . .	73
4.11	Color breakup, average data . . . . .	74
4.12	Effect of stimulus width on color breakup . . . . .	75
4.13	Properties of the impulse-response function . . . . .	76
4.14	Simultaneous vs. alternating capture . . . . .	78

## Acknowledgments

Thank you to Marty Banks, my advisor, for your support throughout graduate school and your endless wisdom on all things vision and displays. Thanks to Joohwan Kim, with whom I collaborated on nearly everything I did throughout my PhD and taught me most of what I know about psychophysics. Also thanks to Samsung Display, particularly David Hoffman, for helping me learn how to think and write like an engineer. Not to mention the fact that his 2011 paper on temporal presentation protocols was essentially my bible in grad school. I'd also like to thank everyone in the lab who repeatedly volunteered their time to help as subjects for psychophysical experiments; I know it was painful. Especially Emily Cooper, who put up with me when I was a rotation student and was instrumental in inspiring me to enter the field of vision science and join Marty's lab. Aljoscha Smolic, Tunç Aydin, and Steven Poulakos at Disney Research Zurich were fantastic collaborators and also made me a greatly improved computer scientist. And a big thanks to my committee—David Whitney and Christoph Schreiner—for helping me prepare this dissertation.

The majority of this work has been published or is currently being prepared for submission:

### Articles

1. J. Kim, **P.V. Johnson**, and M.S. Banks. (2014) Stereoscopic 3D display with color interlacing improves perceived depth. *Optics Express*, 22(26): 31924-31934.
2. **P.V. Johnson**, J. Kim, and M.S. Banks. (2014) The visibility of color breakup and a means to reduce it. *Journal of Vision*, 14(14): 10.
3. **P.V. Johnson**, J. Kim, D.M. Hoffman, A. Vargas and M.S. Banks. (2015) Motion artifacts on 240Hz OLED Stereoscopic 3D displays. *Journal of the Society for Information Display*. (in press)
4. D.M. Hoffman, **P.V. Johnson**, J. Kim, A. Vargas and M.S. Banks. (2015) 240Hz OLED technology properties that can enable improved image quality. *Journal of the Society for Information Display*. DOI #10.1002/jsid.258.
5. **P.V. Johnson**, J. Kim, and M.S. Banks. (2014) Stereoscopic 3D display technique using spatiotemporal interlacing has improved spatial and temporal properties. (submitted to *Optics Express*)

### Proceedings (first author)

1. **P.V. Johnson**, J. Kim, and M.S. Banks. (2013) A novel stereoscopic display technique that minimizes perceptual artifacts. *Journal of Vision*, 13(9): 1173.



2. **P.V. Johnson**, J. Kim, and M.S. Banks. (2014) A novel stereoscopic display technique with improved spatial and temporal properties. *Proc. SPIE-Intl Soc. Optical Eng., Stereoscopic Displays and Applications XXV*, 9011-13.
3. **P.V. Johnson**, J. Kim, D.M. Hoffman, A. Vargas, and M.S. Banks. (2014) 55.1: Distinguished Paper: Motion artifacts on 240Hz OLED stereoscopic 3D displays. In *SID Symposium Digest of Technical Papers* (Vol. 55, No. 1). Blackwell Publishing Ltd.

# Chapter 1

## Introduction

Stereoscopic 3D (S3D) displays send slightly different images to the two eyes, thereby providing an enhanced sensation of depth relative to conventional displays. However, this comes at a cost; delivering different images to each eye on a single display screen is a non-trivial task. There are a variety of methods to accomplish this task, but each is prone to its own unique set of perceptual artifacts, distortions, or sacrifice of resolution (either temporal or spatial). In this introduction I will discuss some of the perceptual consequences associated with different display techniques. Some of these artifacts are specific to stereoscopic displays, while some (e.g. color breakup, motion artifacts) are applicable to non-stereo displays as well.

I will briefly point out that the term “3D display” is not particularly descriptive. Conventional “2D” displays still provide many depth cues and could be argued as providing a sense of three dimensions. Visual cues such as motion parallax, texture, familiar size, occlusion, and shading all provide us with a sense of depth, even in conventional displays. Stereoscopic 3D provide only *one* extra depth cue compared to traditional displays: stereopsis. These displays present separate content to each eye, and the differences between the images are interpreted by the visual system as depth in a process known as stereopsis Held and Hui, 2011. I will use the term “Stereoscopic 3D”, “S3D”, or “stereo” interchangeably throughout this dissertation, but never “3D” alone.

### 1.1 Existing stereoscopic 3D techniques

S3D displays are becoming increasingly important for a variety of applications. In medicine, they have proven useful for applications such catheterization Held and Hui, 2011; Moll et al., 1998, laparoscopic surgery Banks, Read, and Allison, 2012; Byrn, Schluender, and Divino, 2007; Patel, Ribal, Arya, Nauth-Misir, and Joseph, 2007; Rosenthal et al., 2002; Taffinder, Smith, Huber, Russell, and Darzi, 1999, surgical planning Hu, 2006, and medical imaging Hernandez, Basset, Bremond, and Magnin, 1998; Nelson, Ji, Lee, Bailey, and Pre-torius, 2008. Stereo displays also aid in visualizing complex scientific data Frohlich, Barrass,

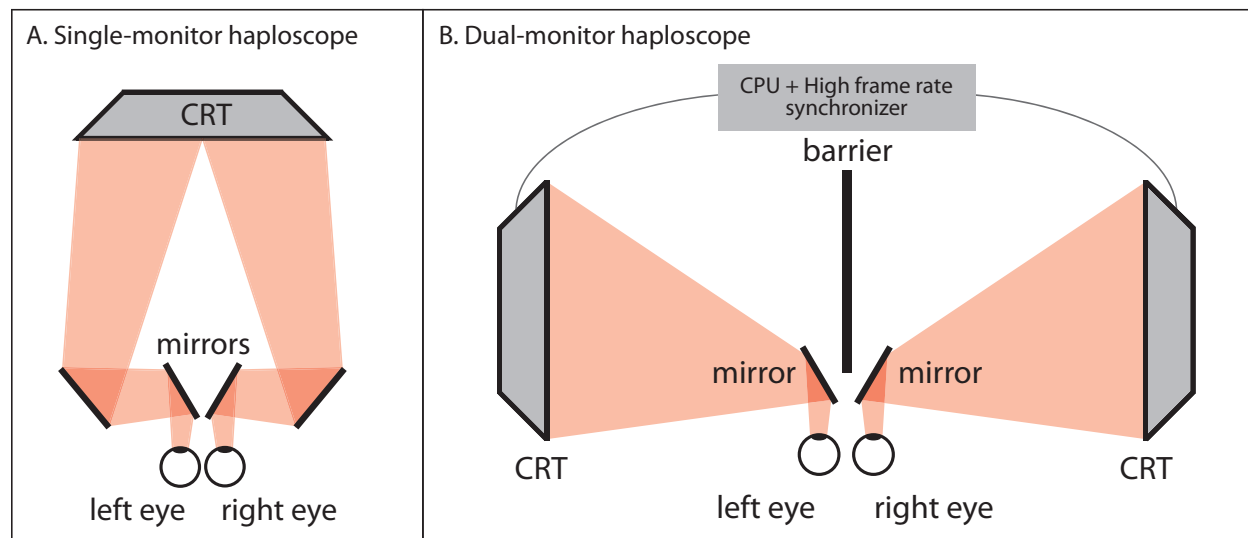


Figure 1.1: Haploscopes used in vision science research. On the left, a single-monitor haploscope, and on the right, a dual-monitor haploscope. These setups allow researchers to have complete control over what images are sent to the left and right eyes, and allow the simulation of different presentation protocols without having to build novel hardware. An additional benefit is that crosstalk—bleeding over of one eye’s image to the other eye—is nonexistent.

Zehner, Plate, and Gobel, 1999 and have clear value in cinema Lipton, 1982 and virtual reality Boman and McMahan, 2007; Chan et al., 2005.

For applications such as cinema and entertainment, misperceptions in S3D displays are, in the worst case, irritating, fatiguing, or nauseating. For applications like surgery and medical imaging, however, perceptual artifacts can make a huge difference in patient outcome and need to be properly understood.

Delivering different images to each eye is the key feature of a stereoscopic display, and the most challenging. The most obvious method involves simply using a separate display screen for each eye, as is the case with a head-mounted display Sutherland, 1968. These displays can provide a large field of view and strong sense of immersion Pastoor, 2005, making them ideal for virtual reality, but have limited use in many applications by the fact that they allow only a single viewer. Another example of a system that uses a dual-screen display is the da Vinci Robot system (Intuitive Surgical, Sunnyvale, CA), which uses an augmented reality stereoscopic visualization for surgery Byrn et al., 2007. Mirror haploscopes are also effective at providing a different screen for each eye (or a different half of a single screen for each eye) and are very useful for vision science research, though they have limited commercial value (see Figure 1.1)Backus, Banks, Van Ee, and Crowell, 1999; Wheatstone, 1838.

Clearly, providing a separate screen for each eye is not always feasible. Commercially available S3D televisions, as well as cinema displays, use a variety of tricks to send different

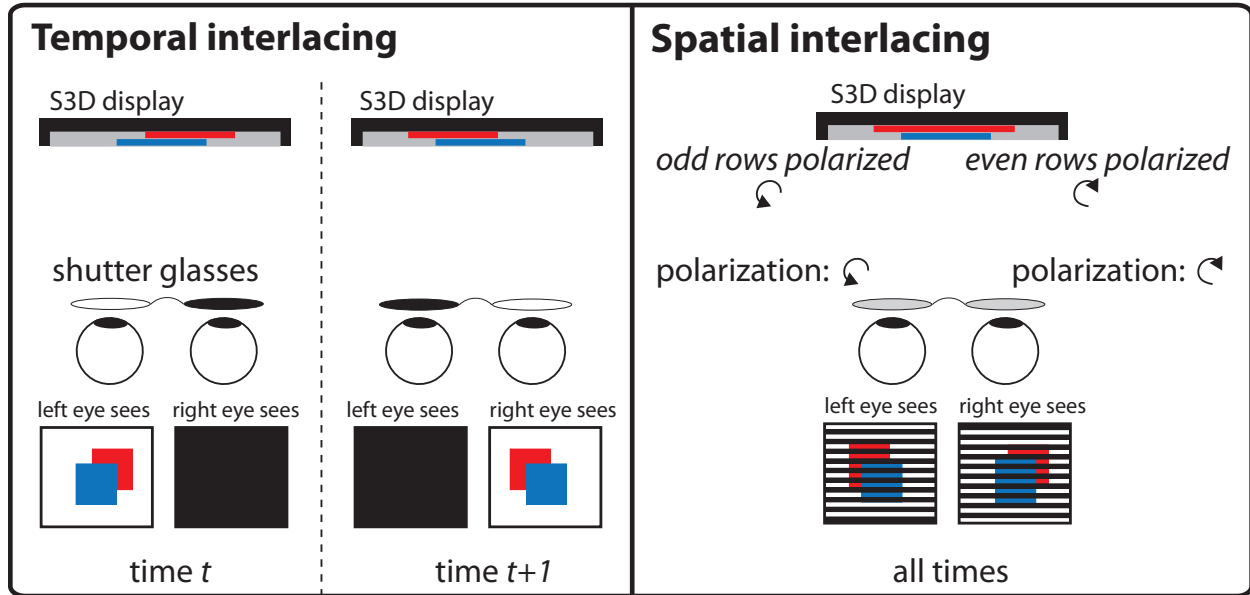


Figure 1.2: Stereoscopic 3D display protocols. Left: Temporal interlacing alternates the left- and right-eye images in time. This is typically achieved using shutter glasses, synchronized with the display, that alternately block the left or right eye to allow only one eye’s view to transmit at a given moment in time. Right: Spatial interlacing sends the even rows to one eye and the odd rows to the other eye. Even rows are polarized clockwise while odd rows are polarized counter-clockwise (or vice versa), and the viewer wears passive polarized glasses to filter out the appropriate view.

images to each eye on a single display screen. Most displays use temporal interlacing or spatial interlacing, illustrated in Figure 1.2.

### Temporal interlacing

*Temporal interlacing* delivers the left- and right-eye views alternately in time. Thus, only one eye receives light at a given moment, but it receives all the pixels. For television, this is often accomplished by using liquid-crystal shutter glasses that alternately transmit and block the images to the eyes in synchrony with the display, often using infrared Dawson, 2012; Mendiburu, 2009. The alternation is rapid enough that the viewer is usually unaware of the fact that one eye’s view is blank at any given moment in time. An alternative approach, used predominantly in cinema, involves projecting polarized images onto a polarization-preserving silver screen Projection of stereoscopic images using linearly polarized light, 2013Mendiburu, 2009. Polarization can either be linear (used by IMAX, Mississauga, Ontario, Canada) or circular (used by RealD, Beverly Hills, CA), and alternates rapidly between different polarization states every frame Sharp and Robinson, 2007. The viewer wears passive eyewear

that filters out the appropriate polarized image at each moment in time. Yet another approach (the Dolby 3D/Infitec method, Ulm, Germany) uses wavelength interlacing, which uses a time-varying color filter that send slightly different narrowband wavelengths of the red, green, and blue channels to each eye Jorke, Simon, and Fritz, 2009; Mendiburu, 2009. The viewer wears passive glasses that transmits the appropriate wavelengths to each eye. Though each eye’s image has a slightly different color gamut, the difference is nearly imperceptible and crosstalk—the contamination of one eye’s image with the other—is lower than other systems Jorke and Fritz, 2003. A company called zSpace (Sunnyvale, CA) also makes a temporally interlaced display that uses a high-speed switchable wave plate to rapidly alternate the polarization state of the display Patterson, 2013. The viewer then wears passive polarized glasses to transmit the appropriate view to each eye.

### Spatial interlacing

*Spatial interlacing* is an alternative approach. This method delivers even pixel rows to one eye and odd pixel rows to the other eye simultaneously Dawson, 2012; Kim and Banks, 2012. This is typically done using a film-patterned retarder on the panel surface that polarizes the emitted light in opposite directions row by row Hong et al., 2011. The polarization can be linear or circular. The viewer wears passive eyewear that transmits even rows for the left eye and odd rows for the right eye (or vice versa). Thus, both eyes receive light at any given moment, but each receives only half the pixels.

### Autostereoscopic displays

Autostereoscopic displays, the holy grail of the 3D TV industry, have made some headway in recent years. *Autostereoscopic* implies that no glasses are required to achieve a stereoscopic 3D effect. Most autostereoscopic displays use either a parallax barrier or a lenticular lens array. Lenticular displays use an array of lenslets oriented vertically, ensuring that each column of pixels is visible only in a certain region of space Holliman, Dodgson, Favalora, and Pockett, 2011. However, because the horizontal resolution of the underlying display is shared between multiple views, the horizontal resolution suffers in proportion to the number of different views Holliman et al., 2011. Though there has been some work that attempts to get around this limitation (e.g. by slanting the lenslets relative to their underlying pixels Van Berkel and Clarke, 1997, this remains a fundamental constraint of this type of display. Parallax barrier displays use vertical opaque columns over the pixel array that enable the observer to see only a subset of the pixels behind the barrier Holliman et al., 2011. This technique has a similar fundamental limitation as lenticular displays, and they are considerably less bright due to the opaque columns that block a substantial portion of the light emitted from the underlying display. Both of these methods also require that the viewer be at a precise location, otherwise artifacts such as image flipping (the discrete transition from one view to a neighboring view) can occur Lambooi, Fortuin, Heynderickx, and IJsselstein, 2009. I won’t discuss autostereoscopic displays much further in this dissertation. They have

a very different set of constraints and engineering hurdles that have thus far prevented them from entering mainstream markets in a big way. There is great potential in these displays, but they likely require a breakthrough in pixel density or panel frame rates before they can present images with adequate resolution and brightness to be competitive against other stereo methods.

### Summary of methods

All these methods have different shortcomings from a perceptual standpoint. Temporal interlacing is prone to temporal artifacts such as flicker, unsmooth motion appearance, and distortions of perceived depth Hoffman, Karasev, and Banks, 2011. Spatial interlacing results in lower spatial resolution at typical viewing distances Kim and Banks, 2012; Park, Kim, and Choi, 2014. Most autostereoscopic displays must make a tradeoff between spatial resolution and number of views. Furthermore, none of these display techniques can provide correct focal cues. I will now discuss in greater detail the perceptual consequences of these display techniques and the mechanisms by which these artifacts occur.

## 1.2 Motion artifacts and flicker

A central pillar of display design is that motion looks smooth only if the display has a sufficiently high frame rate Burr, Ross, and Morrone, 1986. The majority of liquid-crystal displays (LCDs) and organic light-emitting diode displays (OLED) on the market utilize frame rates of 60 frames per second (Hz), producing little flicker and relatively smooth apparent motion. However, there is clear theoretical and empirical evidence that higher frame rates are needed to produce smooth motion for the gamut of typical object speeds Bex, Edgar, and Smith, 1995; Heesch and Klompenhouwer, 2008; Hoffman et al., 2011; Kuroki, 2012; A. B. Watson, 2013; Watson, Ahumada, and Farrell, 1986.

Perceptual motion and flicker artifacts on display systems are influenced by the capture rate and presentation rate. Capture rate is the number of unique images presented per second and is primarily an attribute of the content. Presentation rate is the number of images presented on the screen per second, regardless of whether those images are unique or repeated (multi-flashed), and is limited by the display technology. Capture rate tends to be the primary factor determining the visibility of motion artifacts while presentation rate is the primary factor determining the visibility of flicker Hoffman et al., 2011; Johnson, Kim, Hoffman, Vargas, and Banks, 2015, in press. Bex and colleagues Bex et al., 1995 showed that there is a fixed spatial displacement between image updates that acts as a threshold beyond which temporal aliasing occurs and that this coincides with the point at which motion-energy detection fails Adelson and Bergen, 1985. Additionally, the duty cycle of the image presentation—the fraction of the presentation interval in which imagery is illuminated—affects the visibility of motion artifacts and flicker Hoffman et al., 2011; Watson et al., 1986.

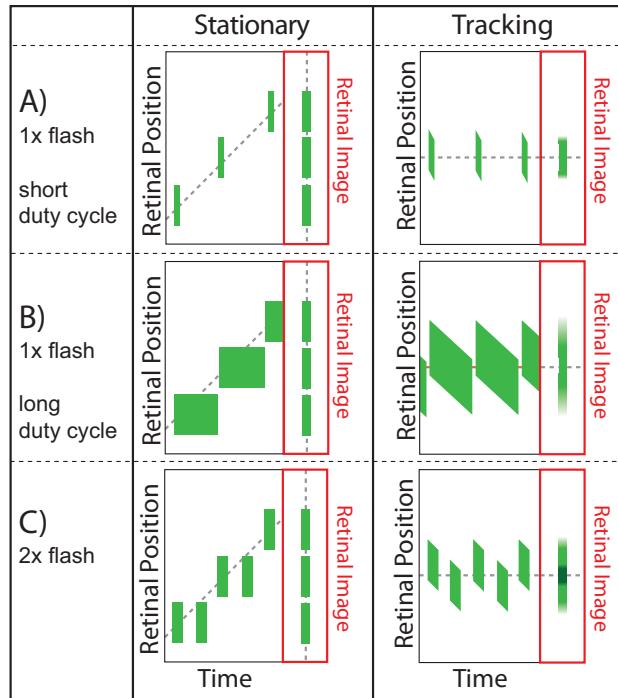


Figure 1.3: Retinal-image stimulation with different display protocols, with stationary fixation and eye tracking. The left sub-region of each panel shows a time and position plot, and the right region shows a cross-section of the retinal image integrated over time. The left panels show the motion along the retina over time when fixation is stationary. The right panels show the retinal motion when the object is tracked with a smooth-pursuit eye movement. A). Single flash (1x), short duty cycle (as in a stroboscopic display). B). Single flash, long duty cycle  $\sim 1.0$  (as in a sample-and-hold display). C). Double flash (2x), duty cycle  $\sim 0.5$  (similar to a temporally interlaced S3D display).

Figure 1.3 summarizes how particular driving modes and viewing conditions stimulate the retina leading to different types of motion artifacts. Consider a viewer fixating on a stationary point on the screen while an object moves past. Because movement on the display is quantized, the object jumps across the retina in discrete steps (Figure 1.3, left column). The displacement of each jump on the retina is the object speed divided by the capture rate of the content. If the displacement is too large, motion appears unsmooth. The unsmooth appearance is called judder. Duty cycle (panels A and B) as well as multiple-flash presentation (panel C) does not impact the spatial position of the retinal image during fixation. Now consider the situation in which the viewer tracks a moving object by making a smooth-pursuit eye movement. With real objects, such tracking stabilizes the object's image on the retina. With digitally displayed objects, the tracking has a different effect, as illustrated in the right column of Figure 1.3. The eye movement causes the discrete image

to smear across the retina for the duration of the presentation interval; this is perceived as motion blur A. B. Watson, 2013. The magnitude of the blur is proportional to the duration of each image presentation and thus motion blur should be greater with longer duty cycles (panel B vs. panel A). Cathode-ray-tube (CRT) displays have an impulse-like temporal response, similar to panel A in Figure 1.3, which keeps motion blur to a minimum. LCDs as well as OLEDs have a sample-and-hold temporal response, similar to panel B in Figure 1.3, suggesting that motion blur could be more prominent in these displays. In cases of multi-flash presentations, another effect—edge banding—can occur (Figure 1.3, panel C) in which repeated presentation of an edge creates the appearance of ghost edges.

Motion artifacts are a spatiotemporal phenomenon involving position and time whereas flicker is purely a temporal artifact. Flicker refers to the sensation of brightness instability. When the duty cycle of a display is less than 1.0, the luminance of a scene shown by the display changes over time. This change becomes visible when the presentation rate is below the critical flicker fusion frequency, which limits the maximum perceptible frequency of luminance change.

The concept of the window of visibility was first proposed by Watson and colleagues and is a simplified band-pass illustration of the visual system and stimulation in Fourier space Watson et al., 1986. It can be used to make predictions of the visibility of different motion artifacts and flicker. Consider an object moving across the screen at speed  $s$  in Figure 1.4. The gray diagonal lines in the left panels represent continuous motion and the blue dots represent stroboscopic sampling of this motion. The Fourier transform of the smoothly moving stimulus is the gray line in the right panels, which has slope  $-1/s$ . Sampling the continuous motion creates replicates: the blue lines. The overall spectrum contains a signal component as well as the replicates. These replicates are only visible if they appear within the window of visibility (schematized by the dashed diamonds in the right panels). This is the region in Fourier space corresponding to the range of spatial and temporal frequencies to which the human visual system is sensitive Watson et al., 1986. The vertex of the window on the temporal-frequency axis represents the critical flicker fusion frequency, or the temporal frequency above which flicker cannot be perceived. Below the critical flicker fusion frequency, flicker visibility will depend on the contrast of the stimulus, with higher contrast stimuli having more visible flicker A. B. Watson, 2013. The vertex of the window on the spatial-frequency axis represents the visual-acuity limit, or the highest spatial frequency that is visible. If aliases are present within the window of visibility, motion artifacts may be visible. The horizontal distance between aliases in Fourier space is equal to  $1/\Delta t$ , where  $\Delta t$  is the rate at which content is captured, suggesting that a higher capture rate would spread aliases further apart, which would make them less likely to infringe on the window of visibility. A capture rate of 60Hz (Figure 1.4, top panels) could cause motion artifacts at this particular object speed, while a capture rate of 120Hz (Figure 1.4, bottom panels) would not. Additionally, the slope of the aliases is the negative reciprocal of speed, so even a capture rate of 120Hz would not prevent motion artifacts at sufficiently high speeds. It should, however, allow for a greater range of speeds that are free of artifacts. Note that if the eyes are tracking the stimulus, we can plot the retinal position over time as a horizontal



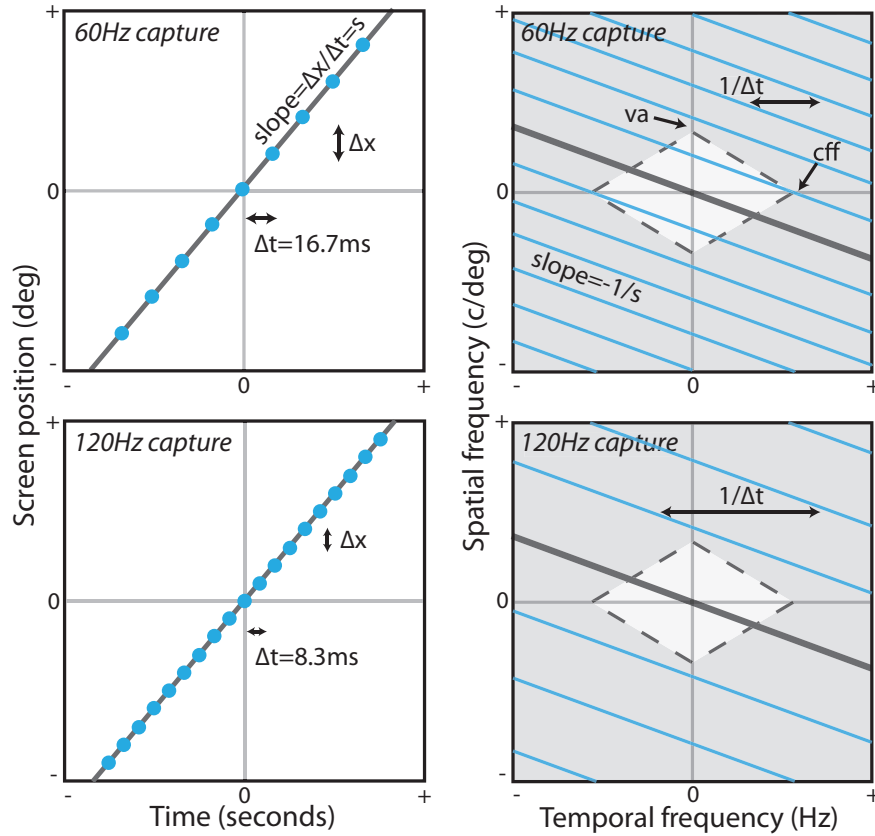


Figure 1.4: Effect of stroboscopic sampling on amplitude spectrum. The gray diagonal lines in the left panels represent smooth motion, and the blue dots represent stroboscopic sampling at two different intervals: 60Hz (top) and 120Hz (bottom). The right panels show the resulting amplitude spectra of the continuous signal (gray line) as well as replicates caused by sampling (blue lines). The diamond represents the window of visibility, the range of spatial and temporal frequencies that is visible. The critical flicker frequency (cff) is the highest visible temporal frequency and the visual-acuity limit (va) is the highest visible spatial frequency. Replicates that fall within the window of visibility can cause motion artifacts, while replicates that remain outside the window are invisible.

line, which would make the signal (and aliases) vertical lines in frequency space. These spatiotemporal aliases would create a different motion artifact percept than when the eyes are stationary Hoffman et al., 2011. This will be discussed further in Chapter 3.

Although presentation rate is an important determinant of flicker visibility Hoffman et al., 2011, other factors—e.g. luminance, contrast, temporal vs. spatial interlacing, duty cycle—contribute as well. Flicker visibility is well predicted by the amplitude and frequency of the Fourier fundamental of the luminance-varying monocular signal from a display. Temporally interlaced S3D displays require duty cycles of 0.5 or less, which increases the amplitude of the fundamental frequency compared to the larger duty cycle on spatially interlaced displays Campbell and Robson, 1968. One therefore expects more visible flicker with temporally interlaced displays compared to spatially interlaced displays. Furthermore, a presentation rate of 60Hz may be inadequate to completely avoid flicker for certain driving modes. There are areas of the peripheral visual field with flicker fusion frequencies as high as 85Hz, while the fovea is  $\sim 55$ Hz Tyler, 1985. This suggests that 60Hz may be sufficiently fast for foveal targets but not for areas in peripheral vision.

### 1.3 Depth distortion

In temporal interlacing S3D displays, the left- and right-eye views are presented in alternation. This means that the second eye sees an image later than the first eye even though the two eyes' contents were captured at the same time. When there is movement in the scene, the visual system interprets the temporal lag as a spatial disparity and perceived depth becomes distorted, a phenomenon known as the Mach-Dvořák effect Burr and Ross, 1979; Hoffman et al., 2011; Read and Cumming, 2005. Consider an object moving horizontally and presented on a temporal-interlaced display (left panel, Figure 1.5). The position of the object is captured with left and right cameras simultaneously at the times marked by black arrows. When the images are presented in alternation, the right image is delayed. The visual system has to match left- and right-eye images to compute disparity, but it is unclear how to make matches because none of the images occur at the same time. If a given left-eye image were matched with the subsequent right-eye image (green arrow in left panel of Figure 1.5), the estimated spatial disparity would be correct (green dots in right panel). But if that same left-eye image were matched with the preceding right-eye image (purple arrow in left panel), the estimated disparity would be incorrect (purple dots in right panel). The brain has no way to know which match is correct because they both have the same inter-ocular time difference. The most reasonable strategy then is to average the two estimates creating a disparity estimate halfway in-between Hoffman et al., 2011; Kane, Guan, and Banks, 2014; Read and Cumming, 2005. The induced disparity,  $\Delta$ , is:

$$\Delta = s\delta \tag{1.1}$$

where  $s$  is object speed and  $\delta$  is the inter-ocular offset of successive presentations. At most frame rates, the viewer does in fact perceive the moving object to be at a depth consistent

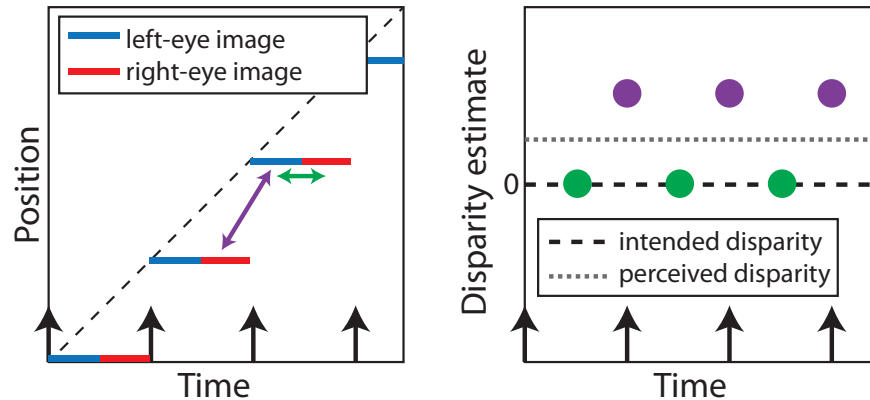


Figure 1.5: Depth distortion in temporally interlaced displays. Left: Temporal interlacing stereoscopic 3D presentation. Right- and left-eye images are captured simultaneously but displayed in alternation. Right: The visual system attempts to match left- and right-eye images, but there is inherent ambiguity in the match, resulting in false matches 50% of the time. The perceived disparity is halfway between the correct matches (green) and false matches (magenta).

with the average disparity estimate Hoffman et al., 2011; Read and Cumming, 2005. For a rightward-moving stimulus with the left-eye image presented before the right-eye image (as in the left panel of Figure 1.5), the time-average estimate is shifted toward crossed (near) disparity, so the object is perceived as closer than intended. For a leftward-moving stimulus, the time-average estimate is shifted toward uncrossed (far) disparity, so the object is seen as farther than intended. This type of depth distortion should not occur with spatial interlacing displays because they present content simultaneously to the two eyes, such that there is no ambiguity about which image in the right eye to match with a given image in the left eye.

There is an artifact known as the Pulfrich effect that has a similar mechanism. Place a neutral-density filter over one eye and view a moving target, and it will appear displaced in depth Morgan and Thompson, 1975. Because the stimulus has a lower luminance in one eye, the neural signal is delayed compared to the eye with greater luminance. This temporal delay is interpreted as a spatial disparity, and the object is seen at the incorrect depth Anzai, Ohzawa, and Freeman, 2001; Morgan and Thompson, 1975. In both the Mach-Dvořák effect and the Pulfrich effect, the depth perceived is dependent on the direction of the motion; motion in one direction will induce a crossed disparity (closer than fixation), while motion in the other direction will induce an uncrossed disparity (further from fixation). The motion of a pendulum swinging back and forth from left to right would therefore appear elliptical in depth.

Any protocol that is able to present the left- and right-eye views simultaneously should have minimal depth distortion of this variety, under the assumption that content is captured simultaneously. If, on the other hand, content is captured alternately, whereby the left- and

right-eye views are captured at the appropriate timestamp for when each eye is presented, the result is different. In this case, temporally interlaced displays should have little depth distortion Hoffman et al., 2011 and spatially interlaced displays should have substantial depth distortion. I will show further evidence of this in Chapter 3.

Content is displayed simultaneously in the spatial interlacing protocol. There is therefore no ambiguity as to how the visual system will match left- and right-eye views and this type of depth distortion should be minimal in this protocol. Another type of depth distortion occurs in spatially interlaced displays when the viewer is close enough to the display to resolve the pixel rows. Because the left- and right-eye views are offset vertically by one pixel, the eyes make a vertical vergence eye movement to binocularly fuse the rows (bright rows aligned in the two eyes and dark rows also aligned). The vertical eye movement causes a change in the horizontal disparity at the retinas of off-vertical and off-horizontal edges, so those edges appear at unintended depths Hakala, Oittinen, and Häkkinen, 2015, in press. Interestingly, some spatial interlacing displays eliminate this effect by presenting data rows alternately. Odd rows on the display are seen by one eye and even rows by the other. But the data presented to odd rows alternate between odd and even and the data presented to even rows alternate between even and odd. The alternation rate is sufficiently high for the alternating data to be temporally averaged by the visual system. This vertical-averaging algorithm eliminates the depth distortion Hakala et al., 2015, in press, but at the cost of reduced spatial resolution Kim and Banks, 2012.

## 1.4 Spatial resolution

Previous studies have shown that spatial interlacing can result in a lower effective resolution than temporal interlacing Kim and Banks, 2012; Park et al., 2014. Kim and Banks (2012) found that viewers' ability to discern fine detail was reduced in spatially interlaced displays when the viewing distance was small Kim and Banks, 2012. The recommended viewing distance for an HD display ( $1920 \times 1080$ ) is 3.1 times the picture height "2022: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays", 2012 "709-5: Parameter values for the HDTV standards for production and international programme exchange", 2002, where pixels subtend an angle on the retina of 1 arcmin. At this viewing distance or closer, temporally interlaced displays have improved resolution compared to spatially interlaced displays. When viewing distance is large enough that pixels are indistinguishable in both displays (around 0.5 arcmin Campbell and Green, 1965b, or  $\sim 6.2$  times the picture height with an HD TV), temporally and spatially interlaced displays have comparable resolutions Kim and Banks, 2012. Figure 1.6 shows results adapted from Kim and Banks, 2012 Kim and Banks, 2012. In another study, Yun and colleagues measured contrast sensitivity on temporally and spatially interlaced displays. When viewing distance was 3.1 times picture height, viewers could discern a Nyquist frequency grating on the temporally interlaced display, but not on the spatially interlaced display Yun, Kwak, and Yang, 2013. Kelley (2011) also reported that because black rows are

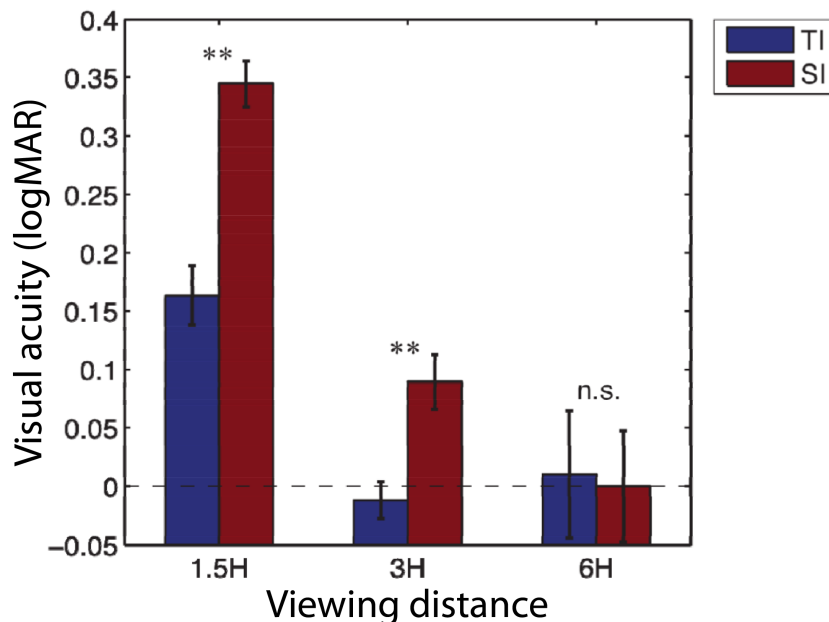


Figure 1.6: Average visual acuity for different interlacing techniques, adapted from Kim and Banks, 2012 Kim and Banks, 2012. Visual acuity (logMAR) is plotted as a function of viewing distance for both spatial and temporal interlacing techniques. Thresholds were estimated using a letter acuity test based on the Bailey-Lovie letter acuity chart Bailey and Lovie, 1976. Letters, of various sizes, were presented on commercially available temporally (or spatially) interlaced televisions, and researchers found the letter size that allowed viewers to respond correctly 75% of the time. When viewing distance was small (1.5H or 3H), temporal interlacing had a better resolution than spatial interlacing. When viewing distance was large (6H), there was no difference between the two interlace techniques.

visible at close viewing distances, viewers will make a vertical vergence eye movement to fuse bright rows, introducing a vertical disparity to the content Kelley, 2011. This disparity has a horizontal component when the contours in the content are neither vertical nor horizontal.

It is worth noting that although the *recommended* viewing distance is 3.1 times the picture height, there is evidence that people tend to prefer sitting further away, depending on the size of the screen Ardito, 1994. However, people prefer sitting closer to their television, relative to screen height, when the television is larger, suggesting that differences between temporal and spatial interlacing may be exacerbated as displays get larger (and our living rooms don't) Ardito, 1994. The move to UHD television ( $3840 \times 2160$ ), however, may shift viewing habits, allowing viewers to sit closer to their displays while keeping pixel size small, which may prove beneficial for spatially interlaced displays. The recommended viewing distance for UHD is 1.55 times the picture height so that the pixel size is 1 arcmin “2020-1: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on

flat panel displays”, 2014. Evidence suggests that viewers prefer to sit closer to UHD displays compared to HD displays Emoto and Sugawara, 2014.

Kim and Banks also sought to determine the difference between binocular and monocular viewing and found that binocular summation provided only a small improvement in effective resolution compared to monocular viewing. This suggests that the limit to effective spatial resolution is determined primarily by the resolution presented to either eye; half-resolution images presented to both eyes do not combine to form a full resolution percept Campbell and Green, 1965a; Kim and Banks, 2012.

## 1.5 Color breakup

Many commercial projectors, such as the popular Digital Light Processing projector (DLP) Hornbeck, 1997, display colors sequentially. The most common implementations present red, green, and blue components sequentially in a given frame. In this case, the presentation rate (the rate at which image data is displayed) is three times the capture rate (the rate at which images are captured by the camera or produced in computer-generated imagery). When such projectors present a bright moving object on a dark background, colored fringes are often seen at the object’s leading and trailing edges Arend, Lubin, Gille, Larimer, and Statler, 1994; Cheng and Shieh, 2009; Post, Monnier, and Calhoun, 1997; Post, Nagy, and Monnier, 1998; Zhang and Farrell, 2003. These fringes are referred to as color breakup.

For a projector presenting red, green, and then blue, the appearance of color breakup depends on the object’s motion and color, and the viewer’s eye movements. When the viewer tracks a moving white object, the three colors land in different places on the retina. Figure 1.7 shows the colored fringes that might be seen with an RGB-sequential protocol when a white object moves rightward across a dark background and the viewer tracks it. Due to the eye movement, the images from later sub-frames are seen displaced leftward relative to those from earlier sub-frames. Thus blue and cyan (green + blue) fringes are seen near the left trailing edge, and red and yellow (red + green) fringes are seen near the right leading edge.

## 1.6 Goals

In these next three chapters, I will present research on novel display technology that attempts to reduce many of these perceptual artifacts. In chapter 2, I discuss a spatiotemporal hybrid display protocol that combines the best properties of spatial and temporal interlacing. In chapter 3, I discuss work that characterized different possible viewing modes of a 240Hz display to determine how to minimize artifacts given the hardware available. Chapter 4 discusses an investigation of color breakup and how to mitigate it.

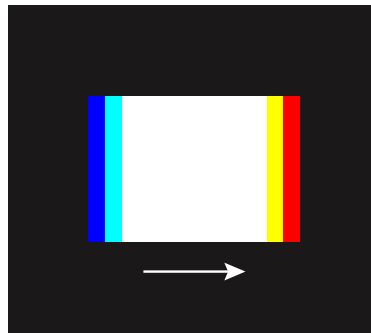


Figure 1.7: Color breakup with an RGB-sequential display protocol when the viewer tracks a moving stimulus. Depiction of a white object moving rightward across a dark background. Red and yellow fringes appear at the leading edge, and blue and cyan fringes at the trailing edge. The width of the fringes depends on the object's speed and display's presentation rate.

## Chapter 2

# Spatiotemporal hybrid display

S3D displays use spatial or temporal interlacing to send different images to the two eyes. Temporal interlacing delivers image to the left and right eyes alternately in time; it is prone to temporal artifacts such as flicker, unsmooth motion, and distortions of perceived depth, but has high effective spatial resolution. Spatial interlacing delivers even pixel rows to one eye and odd rows to the other eye simultaneously; it is subject to spatial limitations such as reduced spatial resolution, but is not prone to temporal artifacts. We propose a spatiotemporal hybrid protocol that interlaces the left- and right-eye views spatially, but with the rows being delivered to each eye alternating in every frame. We hypothesize that this protocol will have the spatial advantages associated with temporal interlacing and the temporal advantages associated with spatial interlacing. To test the hypothesis, we performed psychophysical experiments that compared the temporal and spatial performance of the hybrid protocol to that of the temporal- and spatial-interlacing protocols. We found that spatial resolution in the hybrid protocol is better than in the spatial-interlacing protocol. We also found that flicker, motion artifacts, and depth distortion are reduced relative to the temporal-interlacing protocol. These results suggest that the hybrid protocol retains the benefits of spatial and temporal interlacing while minimizing the drawbacks. The proposed protocol can therefore provide a better viewing experience.

### 2.1 Introduction

We sought a technique that would combine the better features of the two protocols—the spatial resolution of temporal interlacing and temporal performance of spatial interlacing—while minimizing their shortcomings. In the proposed spatiotemporal-interlacing protocol, the left- and right-eye views are interlaced spatially, but the rows presented to each eye alternate temporally. For brevity, we will henceforth refer to the proposed technique as the *hybrid* protocol. Figure 2.1 illustrates the technique.

To describe the protocols we tested, it is useful to define some terms clearly. A *display frame* is the minimal time during which the assignment of a pixel value is maintained. A



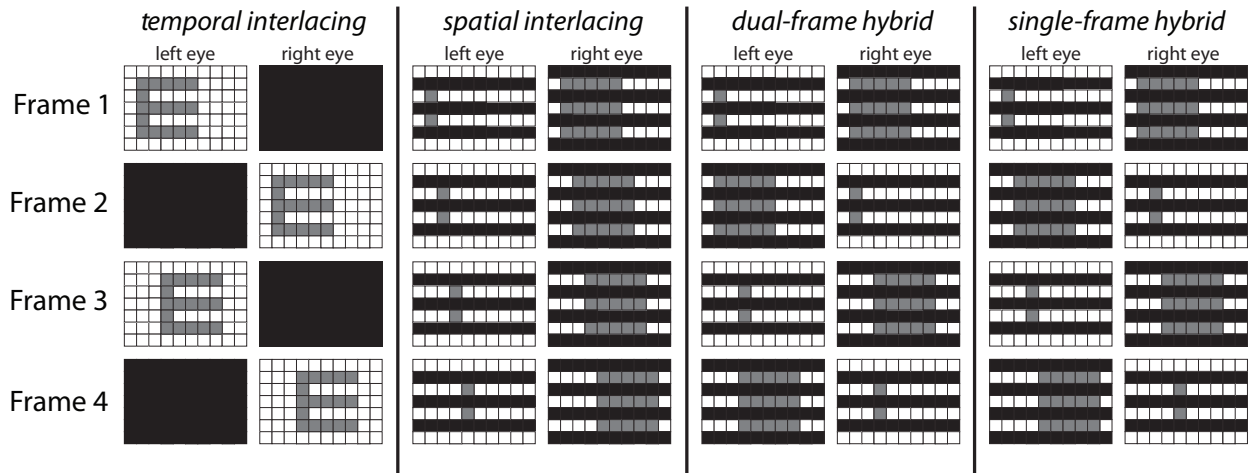


Figure 2.1: S3D display protocols. From left to right, the protocols schematized are temporal interlacing, spatial interlacing, dual-frame hybrid, and single-frame hybrid. To schematize the protocols, we show the images seen by the left and right eyes in two columns for each protocol. Time proceeds from top to bottom. The grid pattern in each panel represents pixels. The stimulus being displayed is a black letter “E” with a height and width of 5 pixels. The stimulus is moving rightward by one pixel per frame such that by frame 5, the E has moved four pixels rightward in all protocols. Black represents pixels that are not displayed to an eye at a given time. In the temporal interlacing and dual-frame hybrid protocols, two display frames are required to show the data captured at one time to both eyes. In the spatial-interlacing and single-frame hybrid protocols, updated image data are shown on every display frame, so the E moves from its previous location with every display frame.

new assignment can occur either to update the image content or to interlace for stereo presentation. Different presentation techniques can require different numbers of display frames to present images to the two eyes. For example, temporal interlacing requires two display frames because it presents one eye’s view at one time and the other eye’s view at another. Spatial interlacing requires only one display frame because it shows both eyes’ views simultaneously. *Display frame rate* is the number of display frames per unit time. *Capture rate* is the number of unique captured (or generated) images per unit time. *Presentation rate* is the number of images (unique or not) presented per unit time. In multi-flash procedures, the presentation rate is the capture rate multiplied by the number of flashes. For example, in the popular triple-flash protocol used by RealD for cinema, the capture rate is 24Hz for each eye but each captured image is displayed three times within a frame for a presentation rate per eye of 72Hz.

There are two possible methods to capture and present content with the hybrid protocol: *dual frame* and *single frame*. In the dual-frame protocol, the captured data are presented

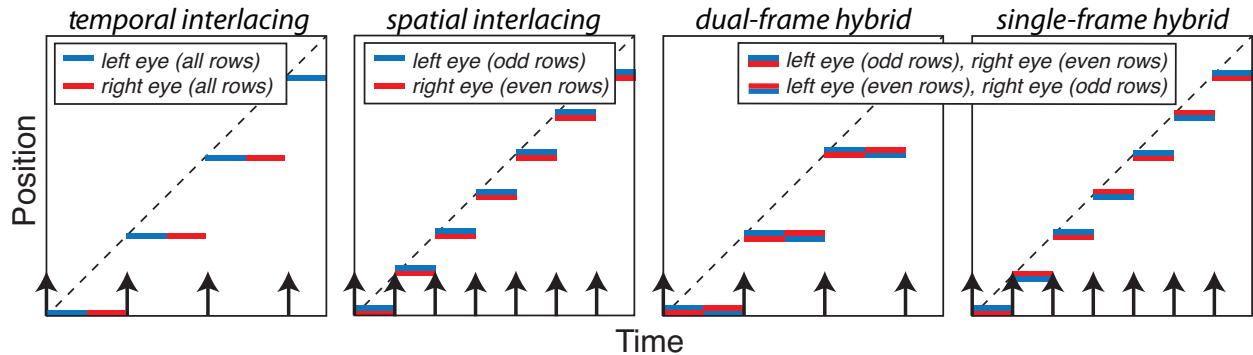


Figure 2.2: S3D display protocols schematized in space-time plots. From left to right, the protocols schematized are temporal interlacing, spatial interlacing, dual-frame hybrid, and single-frame hybrid. Each panel plots position on the screen as a function of time for a stimulus moving at constant speed. The dashed lines represent the objects motion in the real world. The blue and red lines represent the display of the motion on a digital display, blue for images seen by the left eye and red for images seen by the right eye. We assumed a display with a fixed frame rate. Black arrows indicate the times at which content was captured. With a fixed frame-rate display, spatial interlacing and single-frame hybrid allow for presentation of twice the capture rate compared to temporal interlacing and dual-frame hybrid.

over two display frames. In the first display frame, the odd rows of the left eye's image data are displayed in odd rows on the screen and are seen by the left eye, and the even rows of the right eye's data are displayed in even rows on the screen and seen by the right eye. In the second sub-frame, the even rows of the left eye's image data are displayed in even rows on the screen and are seen by the left eye, and the odd rows of the right eye's data are displayed in odd rows and seen by the right eye. Because each display frame presents half the pixel rows of each eye's view, the protocol can present all the captured data. The dual-frame hybrid technique is schematized in the third panel of Figures 2.1 and 2.2. In the single-frame protocol, the captured data are presented on one display frame and updated on every successive frame. In one frame, the odd rows of the left eye's image data are displayed in odd rows on the screen and are seen by the left eye, and the even rows of the right eye's image data are displayed in even rows on the screen and seen by the right eye. In the next frame, new image data are shown, but now the even rows of the left-eye data are displayed in even rows on the screen to be seen by the left eye, and the odd rows of the right eye's data are displayed in odd screen rows to be seen by the right eye. The single-frame hybrid protocol therefore shows only half of the captured data on each frame, but the capture rate is twice that of dual-frame hybrid. This technique is schematized in the fourth panel of Figures 2.1 and 2.2. We compared the four techniques for a fixed display frame rate, so in Figures 2.1 and 2.2 the image data in spatial interlacing and single-frame hybrid are updated

at twice the rate as in temporal interlacing and dual-frame hybrid.

RealD and Samsung developed a presentation technique Sechrist, 2011 similar to the hybrid technique proposed here. Their technique uses two display frames to present S3D image data. Pixel rows are divided into eight blocks across the screen. During the first frame, odd-numbered blocks (1st, 3rd, 5th, and 7th from the top) present the left-eye view and even-numbered blocks (2nd, 4th, 6th, and 8th) the right-eye view. The views swap eyes for the second frame. Hence the difference between the Samsung/RealD technique and the one we propose is how pixel rows get assigned to left- and right-eye views. The Samsung/RealD method spatially alternates between left- and right-eye views every 135 rows (if there are 1080 pixel rows as in HDTV) while our technique does it every other row. At the recommended viewing distance for HDTV, pixel rows subtend 1 arcmin, so the blocks in the Samsung/RealD technique would subtend 135 arcmin, yielding a fundamental spatial frequency of 0.2 cycles/deg (cpd). The blocks in our technique subtend 1 arcmin for a fundamental frequency of 30 cpd. The visual system is much more sensitive to spatial-temporal variations at 0.2 cpd than to such variations at 30 cpd Kelly, 1977, so our technique should provide substantially better image quality and substantially fewer temporal artifacts than the Samsung/RealD technique.

We investigated motion artifacts, flicker, spatial resolution, and depth distortions in four protocols: temporal interlacing, spatial interlacing, dual-frame hybrid, and single-frame hybrid. We found that the proposed hybrid protocol—specifically, the single-frame protocol—has the better properties of temporal and spatial interlacing. Specifically, it has the effective spatial resolution of a temporal-interlaced display while avoiding flicker, motion artifacts, and depth distortions that occur with temporal interlacing.

## 2.2 Experiment 1: Motion artifacts

In the first experiment, we determined the visibility of motion artifacts for the four display protocols, for different object speeds, and when the viewer held fixation stationary or tracked the object.

### Methods

#### Subjects

Six subjects, ages 22 to 32 years, participated. All had normal or corrected-to-normal visual acuity and stereo acuity. Two were authors; the others were not aware of the experimental hypotheses. In all experiments, appropriate consent and debriefing were done according to the Declarations of Helsinki.

## Apparatus

Psychophysical experiments were carried out on a two-display mirror stereoscope. The displays were CRTs (Iiyama HM204DT). A DATAPixx data-acquisition and graphics toolbox (VPixx Technologies) was used to synchronize the two displays precisely. A software package, SwitchResX ([www.madrau.com](http://www.madrau.com)), was used to control the frame rate and resolution of the displays. The resolution of each display was  $1200 \times 900$  so pixels subtended 1 arcmin at the viewing distance of 115 cm. The CRT frame rate was either 100Hz or 75Hz as needed to simulate different capture and presentation rates. Different duty cycles were simulated by presenting one or more CRT frames.

We simulated low display frame rates by repeating CRT frames within a simulated display frame. For instance, we simulated a display frame rate of 50Hz using a CRT refresh rate of 100Hz and repeating each CRT frame twice before updating. The refresh rate of the CRTs was high, so temporal filtering in early vision should make the stimulus effectively the same as an actual sample-and-hold presentation. We checked this by conducting a simulation. We first measured the impulse-response function of the CRTs. We then created sequences of impulses used to simulate the sample-and-hold presentations and convolved them with the temporal impulse-response function of the human visual system Stork and Falk, 1987. The stimulus was a uniform white field. In the simulation, we calculated the temporal modulation of luminance for the CRT and sample-and-hold display. Figure 2.3 plots those modulations as a function of display frame rate and shows that they were very similar. Thus, our means of simulating a sample-and-hold display was valid.

## Stimulus and procedure

We used the stereoscope to send the appropriate content to each eye at each moment in time. By so doing, we could simulate the four S3D display protocols (Figures 2.1 and 2.2). We measured the visibility of motion artifacts by presenting a series of moving  $1^\circ$  bright squares separated by  $3^\circ$  on an otherwise dark background (Figure 2.4). Stimulus duration was 1 sec. Stimuli were presented binocularly with zero disparity. We used MATLAB with the Psychophysics Toolbox extension to render and display all content Brainard, 1997; Pelli, 1997. We adjusted luminances so that stimulus contrast was equivalent in the four display protocols. We presented two eye-movement conditions: a tracking condition, in which subjects made smooth movements to track the stimulus, and a non-tracking condition, in which fixation was stationary as the stimulus moved by. In the tracking condition, a fixation cross was presented to one side 0.5 sec before the stimulus appeared. The cross moved across screen with the stimulus to aid tracking (Figure 2.4, left). In the non-tracking condition, a fixation cross was presented at screen center 0.5 sec before the onset of the stimulus. Then the stimulus moved adjacent to the stationary cross (Figure 2.4, right). Stimulus motion was either horizontal (left to right, or right to left) or vertical (top to bottom, or bottom to top).

Subjects indicated after each trial whether they had seen motion artifacts or not, regard-

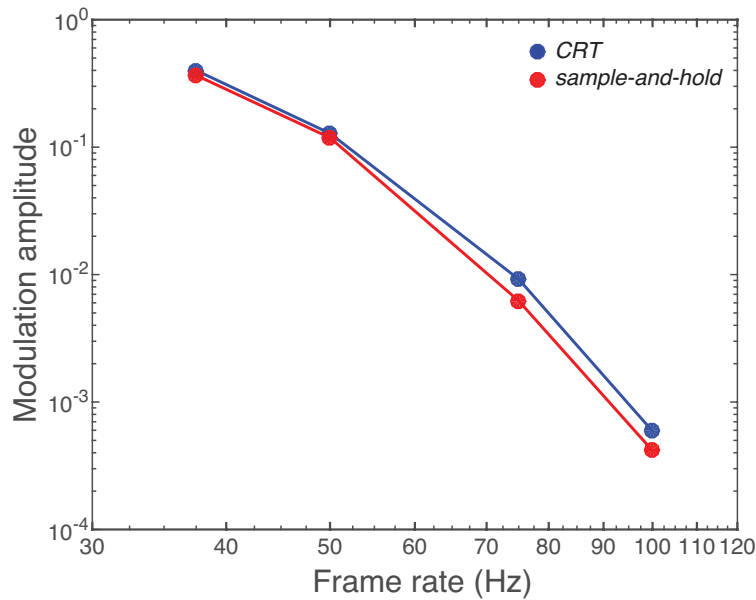


Figure 2.3: Calculated luminance modulation for CRT and sample-and-hold displays. Peak-to-trough luminance modulation after filtering by the human temporal impulse-response function is plotted as a function of display frame rate. Blue represents modulation for the CRTs and red represents modulation for a sample-and-hold display with instantaneous on and off responses.

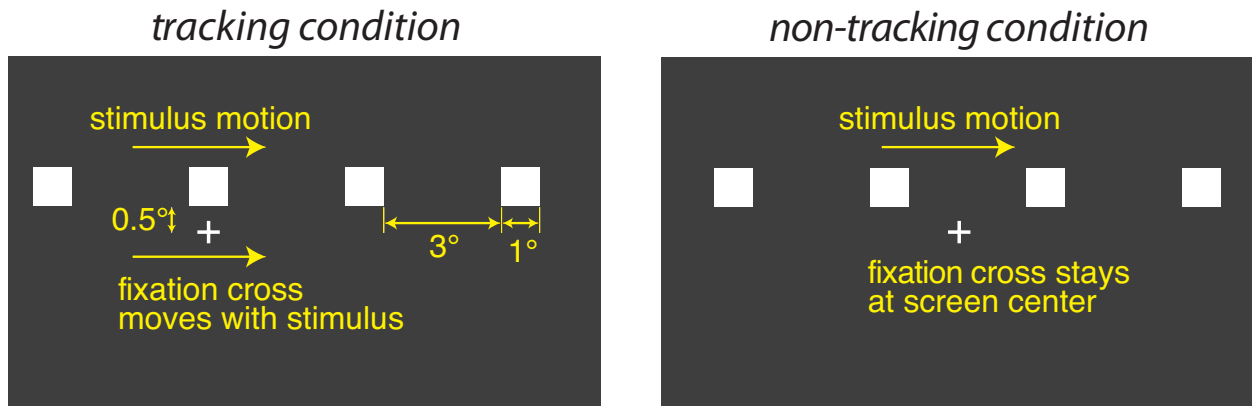


Figure 2.4: Stimulus used to measure visibility of motion artifacts. In the tracking condition, the fixation target moved with the same velocity as the squares. In the non-tracking condition, the fixation target remained stationary as the squares moved by. In both cases, the fixation target appeared 0.5 sec before stimulus onset.

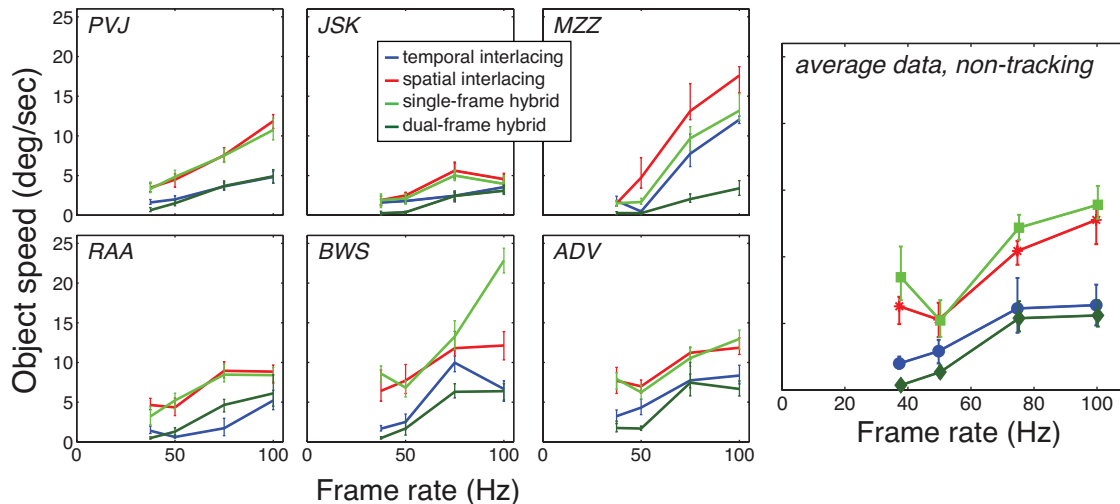


Figure 2.5: Visibility of motion artifacts for the four protocols in the non-tracking condition. The six panels on the left show the data from individual observers. The large panel on the right shows the data averaged across observers. Each panel plots the stimulus speed at which artifacts were reported on half the trials as a function of display frame rate. Blue, red, bright green, and dark green represent the results for temporal interlacing, spatial interlacing, single-frame hybrid, and dual-frame hybrid, respectively. Error bars represent 95% confidence intervals. Temporal interlacing and dual-frame hybrid require two display frames per capture period, while spatial interlacing and single-frame hybrid require only one frame. For a given display frame rate, spatial interlacing and single-frame hybrid can therefore present twice the capture rate, allowing for smoother motion appearance.

less of the type of artifact (e.g., edge banding, blur, or judder). This was a 2-alternative, forced-choice judgment. A 1-up/1-down adaptive staircase procedure adjusted the speed of the stimulus to estimate the value that just yielded motion artifacts. Twenty trials were presented for each staircase. Staircases were randomly interleaved within an experimental session. Maximum speed was  $20^\circ/\text{sec}$ . We fit the data with a cumulative Gaussian whose parameters were determined with a maximum-likelihood criterion Fründ, Haenel, and Wichmann, 2011; Wichmann and Hill, 2001a, 2001b. Henceforth we report the stimulus speed at which the Gaussian crossed 50%, which is an estimate of the speed at which motion artifacts were perceived on half the trials. When we averaged across subjects, we did so by pooling the psychometric data from all subjects and then fitting those data with one cumulative Gaussian function.

The experiment consisted of  $\sim 1280$  trials per subject: 4 display protocols  $\times$  4 capture rates  $\times$  2 eye-movement conditions  $\times$  2 directions  $\times$  20 trials. It took about 1 hour for each subject to complete the experiment.

## Results

Figure 2.5 plots the data from the non-tracking condition. Each panel shows the object speed at which observers reported motion artifacts on half the trials as a function of display frame rate. Different colors represent the data from different protocols. There were clear differences across observers in the speeds at which they reported artifacts, but they all exhibited the same effects across protocols. The threshold speeds for vertical and horizontal motion were not significantly different for any of the protocols, so we averaged the data across the two motion directions. Of greatest interest is how the different protocols fared in terms of artifact visibility. The results show, as expected, that temporal interlacing is more prone to motion artifacts than spatial interlacing Hoffman et al., 2011; Johnson et al., 2015, in press. Artifacts with the dual-frame hybrid protocol were similar to those with temporal interlacing, while artifacts with the single-frame hybrid protocol were similar to those with spatial interlacing. Thus, the single-frame version of the hybrid technique is relatively immune to motion artifacts.

From previous work, we expect capture rate and object speed to be the primary determinants of motion artifacts Hoffman et al., 2011; Johnson et al., 2015, in press. Specifically, whenever the ratio  $S/R_C$  (where  $S$  is speed and  $R_C$  is capture rate) exceeds a critical value, artifacts should become visible. Figure 2.6 plots the average data in Figure 2.5 as a function of capture rate. Plotted this way, the speed at which artifacts became visible is very similar across protocols. The dashed line is  $S/R_C = 0.136$ , so the data show that artifacts became visible whenever that ratio exceeded 0.136. These results are quite similar to those of Hoffman et al. Hoffman et al., 2011 and Johnson et al. Johnson et al., 2015, in press who observed a critical ratio of 0.1 to 0.2.

The results from the tracking condition are shown in Figure 2.7. As you can see, artifacts became less visible with tracking (that is, higher speeds were required to produce them than in the non-tracking condition) in the spatial interlacing, temporal interlacing, and single-frame hybrid protocols. Observers reported that tracking also changed the type of artifact seen. With tracking, motion blur became the most frequent at higher frame rates; with stationary fixation, most artifacts were judder and edge banding. The visibility of motion blur should depend on how much each stimulus presentation is smeared across the retina: specifically, on how much the stimulus moves across the retina during a presentation. The displacement on the retina is proportional to object speed (because the speed of the eye movement is determined by that speed) and the “on” time of a single presentation: that is,  $D = ST$ , where  $D$  is displacement,  $S$  is speed, and  $T$  is “on” time. We examined whether the product  $ST$  is the determinant of artifact visibility, hence whether motion blur was the primary artifact in the tracking condition. Figure 2.8 re-plots the data from Figure 2.7 as a function of presentation time (the time the stimulus is illuminated for one eye during the presentation of one content frame) in log-log coordinates. If a certain retinal displacement  $D_C$  is required to create visible motion blur, the data should be predicted by the equation  $S = D_C/T$ , which is equivalent to  $\log S \propto -\log T$ . As you can see, this equation provides a good fit to the data with short presentation times, but not with long presentation times. In

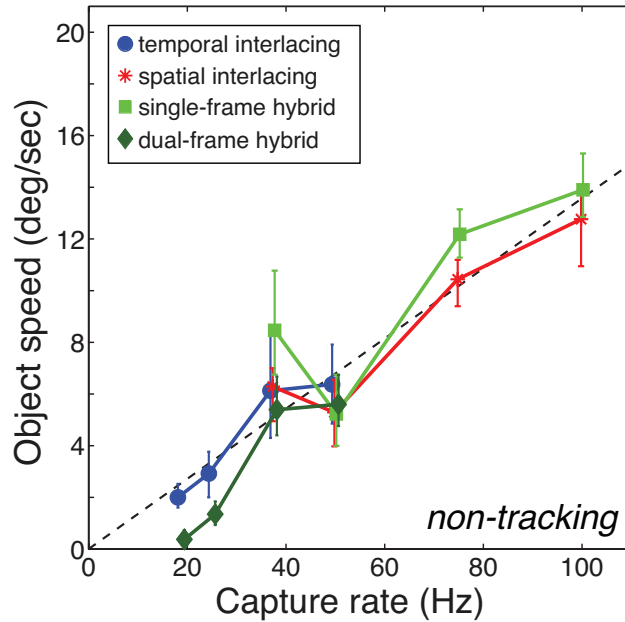


Figure 2.6: Capture rate and the visibility of motion artifacts for the non-tracking condition. The speed at which motion artifacts were reported on half the trials is plotted as a function of capture rate. The data are the same as the right panel in Figure 2.5, but plotted on a different abscissa. Temporal interlacing and dual-frame hybrid require two display frames per capture period, so the maximum possible capture rate was 50Hz for those protocols. Blue, red, bright green, and dark green represent the results for temporal interlacing, spatial interlacing, single-frame hybrid, and dual-frame hybrid, respectively. Error bars represent 95% confidence intervals. The dashed line represents the equation  $S/R_C = 0.136$ .

a follow-up experiment, we asked three of the six observers to indicate on each trial whether they had seen motion blur or not. The open circles represent those data, which were indeed well predicted by the equation  $S = D_C/T$ . This confirms that motion blur is determined by how much a single presentation displaces across the retina. We think the additional artifacts seen at longer presentation times (filled symbols) are caused by small eye movements that produce apparent flicker.

We conclude that motion artifacts in the single-frame hybrid protocol are no more visible than in the spatial-interlacing protocol whether the viewer is tracking the stimulus or not. Therefore, the single-frame version of the hybrid protocol is as immune to motion artifacts as conventional spatial interlacing and is more immune than temporal interlacing.



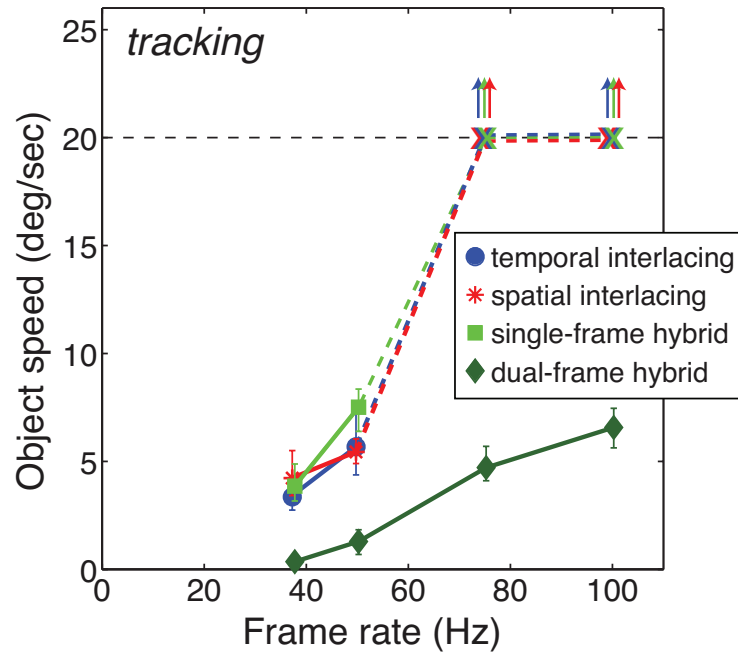


Figure 2.7: Visibility of motion artifacts with tracking eye movements. The speed at which motion artifacts were reported on half the trials is plotted as a function of display frame rate. Data have been averaged across subjects. Different colors represent the data from different protocols. Error bars represent 95% confidence intervals. The horizontal dashed line represents the maximum speed tested, so data points (X's) plotted on that line indicate conditions in which no artifacts were reported at any tested speed.

## 2.3 Experiment 2: Flicker

### Theory

Visible flicker is defined as perceived fluctuations in the brightness of a stimulus due to the digital display of the stimulus. Presentation rate has been shown to be the major determinant of flicker visibility Hoffman et al., 2011; Hoffman, Johnson, Kim, Vargas, and Banks, 2015, in press; Johnson et al., 2015, in press. The threshold for temporally interlaced S3D displays is  $\sim 40$ Hz Hoffman et al., 2011. The threshold value is well predicted by the amplitude and frequency of the Fourier fundamental of the luminance-varying signal from the display Cavonius, 1979; Hoffman et al., 2011; Watson et al., 1986. The temporal frequency of the Fourier fundamental differs across protocols when the display frame is the same. For temporal interlacing, the fundamental frequency is half the display frame rate because the display has to alternate between the two-eyes' views. For the two hybrid protocols,

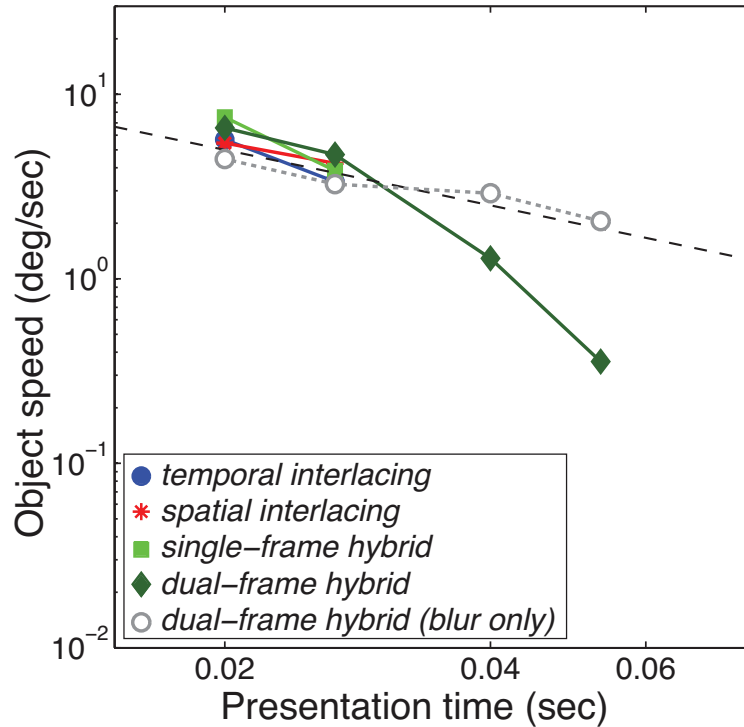


Figure 2.8: Motion artifacts and presentation time. The speed at which motion artifacts were reported on half the trials is plotted as a function of presentation time. The solid symbols and lines represent the data from Figure 2.7 with the exception of conditions in which no artifacts were reported. The open symbols and dotted line represents data in the follow-up experiment in which observers were asked to report motion blur only. The dashed line represents the prediction that motion blur is seen whenever displacement across the retina exceeds a critical value:  $ST > D_C$ .

it is also half the display frame rate, but the phase is shifted between the even and odd pixel rows. For spatial interlacing the fundamental frequency is the same as the display frame rate. Duty cycle (the time the stimulus is illuminated in one eye divided by the “on” plus “off” time in the same eye) also plays an important role. Shorter duty cycles create greater amplitudes of the fundamental frequency, so one expects flicker to be more visible as duty cycle is decreased. Temporal-interlacing displays cannot present duty cycles greater than 0.5 because each eye receives a black frame for at least half the time while the display sends light to the other eye. Spatial-interlacing displays, on the other hand, can have duty cycles as great as 1 because content is presented simultaneously to both eyes. In the hybrid protocols, duty cycle is 0.5 for each pixel row. Spatial interlacing should be the most immune to flicker because it has a higher fundamental frequency and a longer duty cycle than the other protocols. Temporal interlacing and the two hybrid protocols have the same

fundamental frequency and duty cycle, but their spatial characteristics are different. Any image presented using only odd (or even) pixel rows has dark stripes on even (or odd) pixel rows. Thus the spatial frequency associated with the temporal alternation is much higher than with temporal interlacing. Flicker is less visible at high than at low spatial frequencies so flicker should be much less visible in the hybrid and spatial-interlacing methods than in the temporal-interlacing protocol.

## Methods

### Subjects

Four subjects, ages 22 to 32, participated in the experiment. All had normal or corrected-to-normal visual acuity and stereo acuity. Two were authors; the rest were not aware of the experimental hypotheses.

### Apparatus

The experiments were conducted on one display (ViewSonic G225f) seen binocularly via a mirror stereoscope. SwitchResX was used to control the frame rate and resolution. Viewing distance was 213cm.

### Methods

We determined the display frame rate at which flicker was just visible for each protocol. The stimulus was a stationary bright  $1 \times 1^\circ$  square on a dark background. It was presented binocularly with zero disparity for 1sec. Luminance and contrast were equivalent for the tested protocols. The resolution of the display was  $1066 \times 800$ , yielding a pixel size of 0.6 arcmin. Flicker visibility is strongly dependent on spatial frequency Kelly, 1979. To determine how the spatial frequency of the alternating blocks affected flicker visibility, we varied block height. The heights  $h$  were 0.6, 1.2, 1.8, 2.4, 3.6, 6.0, 8.4, and 14.4 arcmin. These correspond respectively to spatial frequencies per eye of 50, 25, 16.7, 12.5, 8.3, 5, 3.6, and 2.1 cpd. The block height in the Samsung/RealD protocol viewed at the recommended viewing distance for HDTV was 135 arcmin, corresponding to a spatial frequency of 0.2 cpd. The interaction of flicker visibility and spatial frequency is also highly dependent on retinal eccentricity Koenderink, Bouman, de Mesquita, and Slappendel, 1978. To examine the effect of eccentricity, we presented stimuli in two retinal positions: on the fovea and  $4^\circ$  below the fovea. The CRT was able to use various frame rates (80, 90, 100, 120, 130, 140, or 150Hz), allowing us to simulate a large set of display frame rates: 6, 8, 10, 20, 30, 40, 50, 60, 80, 90, 100, 120, 140, and 150Hz. After every trial, subjects indicated whether they had seen flicker or not.

The experiment consisted of  $\sim 6240$  trials per subject: 3 display protocols  $\times$  8 block sizes  $\times$  2 retinal locations  $\times$  13 presentation rates  $\times$  10 trials. About 3 hours were required for each subject to complete the experiment. There was only one hybrid protocol because the

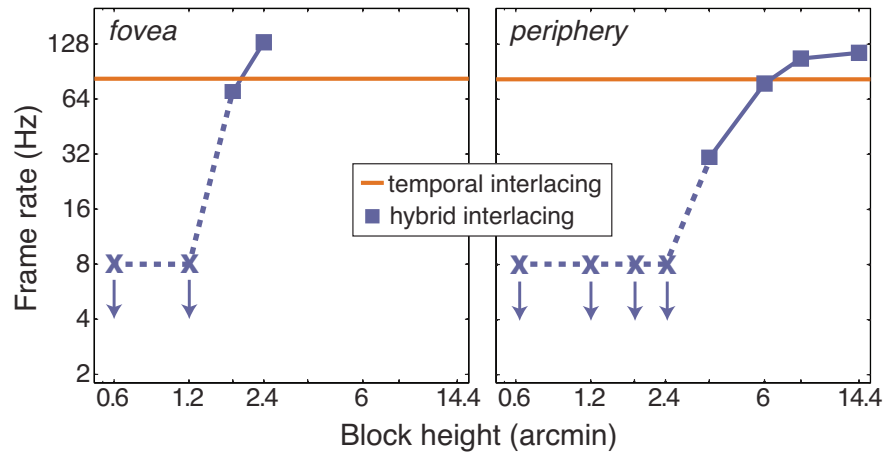


Figure 2.9: Flicker visibility for different protocols and block heights. Each panel shows the display frame rate at which flicker was reported on half the trials as a function of the height of the blocks of pixels that were alternated. The left panel shows the data when the stimulus was on the fovea and the right panel the data when the stimulus was  $4^\circ$  below the fovea. Data have been averaged across observers. Orange and blue represent the data with the temporal-interlacing and hybrid protocols, respectively. Flicker was never visible with the spatial-interlacing protocol even at the lowest tested display frame rate of 8Hz, so we do not plot those data. Flicker was also never visible with the hybrid protocol at the lowest tested display frame rate when blocks were small; those points are represented by X's. Error bars represent 95% confidence intervals; some are too small to be visible.

dual- and single-frame protocols are identical when the stimulus is stationary. We used the method of constant stimuli to vary presentation rate in order to find the rate at which flicker was reported on half the trials. As in Experiment 1, we fit a cumulative Gaussian to the resulting psychometric data using a maximum-likelihood criterion and used the 50% point on that function as the estimate of the rate that produced just-visible flicker. All conditions were randomly interleaved.

## Results

Figure 2.9 plots the display frame rate that produced just-visible flicker as a function of block height. Block height was irrelevant to the temporal- and spatial-interlacing protocols, so there could be only one estimated rate at which flicker was seen for each of those protocols:  $\sim 82$ Hz for temporal interlacing and none for spatial (i.e., flicker was not seen even at the lowest presented rate). The frame rate of 82Hz corresponds to the monocular presentation with a fundamental frequency of 41Hz, which agrees with the previously reported results Hoffman et al., 2011. Thus, as expected, flicker was significantly more visible with temporal

interlacing than with spatial interlacing. The hybrid results are more interesting. With this protocol, blocks are alternately illuminated, so when flicker is perceived, it is at the scale of individual blocks. As we said earlier, sensitivity to high temporal frequencies decreases with increasing spatial frequency Kelly, 1977, and that decrease occurs at lower spatial frequencies in the periphery than in the fovea Koenderink et al., 1978. For these reasons, we expected that flicker would become more visible as block height increased and that the height producing flicker would be greater in the periphery than in the fovea. This is precisely what we observed. In the fovea, flicker was not seen when block height was 0.6 arcmin and then became more visible as height increased from 1.2-2.4 arcmin. In the periphery, flicker was not perceived when block height was 0.6-2.4 arcmin and then became visible at greater heights. Thus, the hybrid protocol is relatively immune to flicker provided that the size of each alternating block of pixels is smaller than 2 arcmin.

The recommended viewing distance for HD television is 3.1 times the picture height and for UHD is 1.55 times picture height “2020-1: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays”, 2014 “2022: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays”, 2012 “709-5: Parameter values for the HDTV standards for production and international programme exchange”, 2002. Both displays at the recommended distances yield pixels subtending 1 arcmin. We observed negligible flicker with the hybrid protocol when block height was less than 2 arcmin, so an implementation of this protocol should produce essentially no visible flicker when viewed at the recommended distances or farther.

It is interesting that the hybrid protocol produced more visible flicker than temporal interlacing when the block heights were 2.4 arcmin and larger in the fovea and 6 arcmin and larger in the periphery. We believe this increased visibility is caused by small eye movements, or microsaccades, that dart back and forth across the boundaries between alternating blocks. Davis, Hsieh, and Lee Davis, Hsieh, and Lee, 2015 demonstrated that these small eye movements can cause visible flicker even when the display’s alternation rate is very high. They divided the screen of a display with a very high frame rate into left and right halves. The luminance of each half was modulated at very high temporal frequencies, but in opposite phases. Subjects saw flicker at the boundary between the two halves, even when the modulation rate was as high as 500Hz. Davis and colleagues argued persuasively that the perceived flicker was due to high-frequency horizontal eye movements across the vertical alternation boundary causing different parts of the retina to be exposed to modulation rates that were occasionally much lower than 500Hz. We believe the same effect underlies flicker visibility with the hybrid approach when pixels are sufficiently large.

From our results, we believe that the Samsung/RealD hybrid display created very noticeable flicker because the heights of the alternating blocks were much greater than the heights for which our subjects reported flicker even at high frame rates.

## 2.4 Experiment 3: Spatial resolution

### Theory

If images are presented in every pixel row to an eye, the spatial frequency due to the rows is 30cpd at the recommended viewing distances for HD-TV and UHD-TV “2020-1: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays”, 2014 “2022: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays”, 2012 “709-5: Parameter values for the HDTV standards for production and international programme exchange”, 2002. At 30cpd, the rows would be barely visible. With spatial interlacing, images are presented in every other row to an eye, so the spatial frequency is 15cpd per eye making the rows more visible monocularly. There are claims that the visual system can fuse two monocular images like those in spatial interlacing to form a binocular image with no missing rows Soneira, 2011; de Witt, 2012. If these binocular-fusion claims are correct, the effective spatial resolution of a spatial-interlacing display would be the same as a temporal-interlacing display that presents all rows to each eye. If these claims are incorrect, however, one would have to double the viewing distance with spatial interlacing to make the rows roughly equally visible compared to temporal interlacing. We measured the spatial resolution of different protocols to see if effective resolution is indeed reduced in spatial interlacing and to see if the hybrid protocols provide greater effective resolution than spatial interlacing.

Kim and Banks Kim and Banks, 2012 measured effective spatial resolution with spatial and temporal interlacing at different viewing distances. They found that viewers’ ability to discern fine detail was reduced with spatial interlacing provided that the viewing distance was not too great. They observed the resolution difference with both monocular and binocular viewing suggesting that the binocular-fusion claim is incorrect. Hybrid interlacing should provide greater effective resolution than spatial interlacing because each eye receives a full-resolution image, albeit over the course of two frames. If no object motion is present, the visual system can average over the two frames to gain high resolution. If motion occurs, however, the gain in resolution may depend on the direction and speed of motion.

The effective spatial resolution of any display is affected by viewing distance. At long distances, where pixels are too small to be resolved by the visual system, resolution becomes “eye limited” Kim and Banks, 2012. The size and arrangement of pixels will therefore not matter to the viewer’s ability to see fine detail. At shorter distances, where pixels can be resolved by the visual system, resolution becomes “display limited” Kim and Banks, 2012 and then the size and arrangement of pixels affects the viewer’s ability to see fine detail.

We examined the effective spatial resolution of the four protocols illustrated in Figures 2.1 and 2.2 and also determined how vertical and horizontal motion influences the outcome.

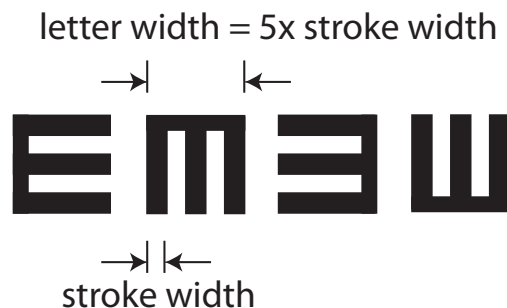


Figure 2.10: Stimuli used to measure spatial resolution. The height and width of the letter was always five times the stroke width. Thus, when letter size was manipulated, the stroke width changed as well as the letter height and width. On each trial, subjects indicated which of the four orientations was presented.

## Methods

The same six subjects participated as in Experiment 1. The one-display stereoscope from Experiment 2 was used. We determined the effective spatial resolution of the four display protocols with a “tumbling E” task. In this task, observers report which of four orientations of the letter E was presented (Figure 2.10). The size of the letter was varied to find the just-identifiable size. The letter was black on an otherwise white background. The stimuli were presented stereoscopically with a disparity of 0. Viewing distance was 213cm and display resolution was  $1280 \times 960$ , which created a pixel size of 0.5 arcmin. At the recommended viewing distances for HD- and UHD-TV, a pixel subtends 1 arcmin.

We simulated changes in viewing distance by simulating pixels of different sizes—one pixel for a simulated pixel size of 0.5 arcmin,  $2 \times 2$  pixels for 1 arcmin, and  $4 \times 4$  for 2 arcmin—and having subjects view from a fixed distance of 213cm. We did this instead of actually changing viewing distance so that we could randomly interleave all experimental conditions. We verified that our method for simulating the effect of viewing distance was valid by conducting a control experiment. In the control experiment, we used the same Tumbling-E task to measure letter-acuity thresholds in three participants for the temporal-interlacing, spatial-interlacing, and hybrid protocols. The letters did not move so the two versions of the hybrid protocol were identical. We measured acuity when the viewing distance was actually varied (distances of 213, 106, and 53cm;  $1 \times 1$  pixels) and when changes in distance were simulated (distance was fixed at 213cm, and pixels were  $1 \times 1$ ,  $2 \times 2$ , and  $4 \times 4$ ). The measured acuities did not differ systematically (expressed in angular units), which shows that our method of simulating different viewing distances was valid.

Frame rate was 120Hz. We presented static and moving stimuli; motion was vertical or horizontal at a speed of  $3^\circ/\text{sec}$ . On the motion trials, a fixation cross was presented eccentrically 0.5 sec before stimulus onset in order to inform the subject of the direction of the upcoming movement. The stimulus then crossed screen center and the viewer tracked

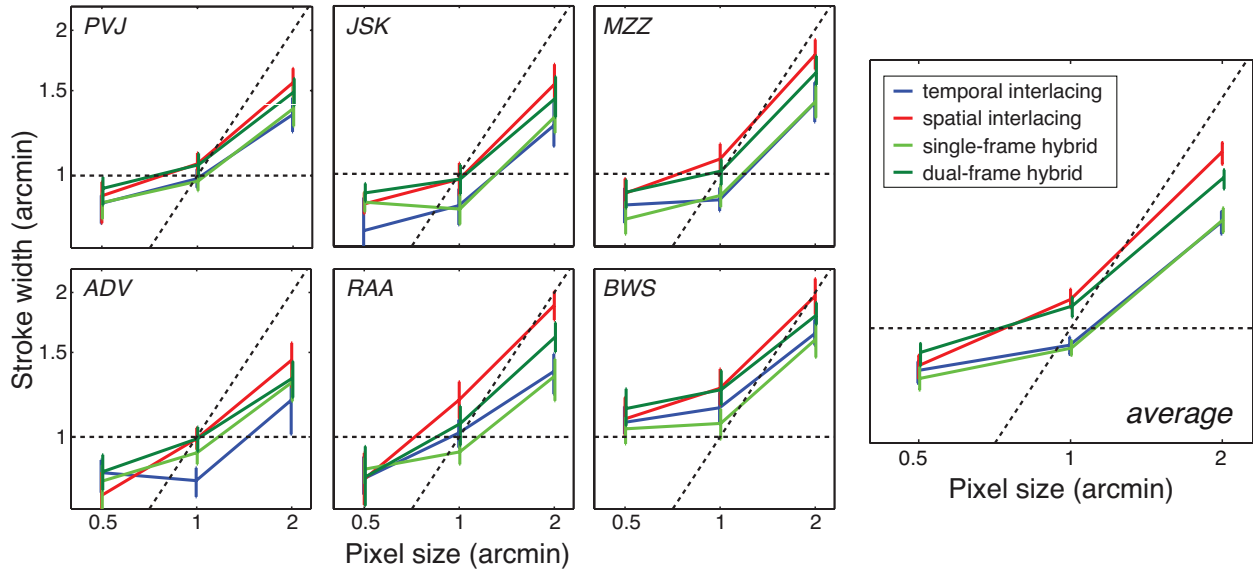


Figure 2.11: Spatial resolution for different protocols and simulated viewing distances. The six panels on the left plot the data from individual observers, averaged across the three motion conditions. Each panel plots the letter stroke width for which the observer identified letter orientation correctly on 62.5% of the trials. Different colors represent the data from different protocols. Error bars represent 95% confidence intervals. The horizontal and diagonal dashed lines represent the expected values for eye-limited and display-limited acuities, respectively, on a conventional 2D display. The right panel shows the data averaged across subjects.

it with their eyes. Stimuli were presented for 0.6 sec and the viewer responded up, down, left, or right to indicate the perceived orientation of the letter. The task was therefore a 4-alternative, forced-choice task. No feedback about the correctness of each response was provided.

The experiment consisted of 10,368 trials per subject: 4 display protocols  $\times$  3 pixel sizes  $\times$  3 movements  $\times$  4 orientations  $\times$  9 letter sizes  $\times$  8 trials. About 3 hours were required for each subject to complete the experiment. We used the method of constant stimuli to vary letter size. We fit the resulting psychometric data with a cumulative Gaussian using a maximum-likelihood criterion. The acuity threshold for each condition was defined as the letter size for which orientation was identified correctly on 62.5% of the trials.

## Results

The results are shown in Figure 2.11. The six panels on the left show the data from the individual observers. The data have been averaged across the three motion conditions. The panel on the right shows those data averaged across observers. The horizontal and diagonal



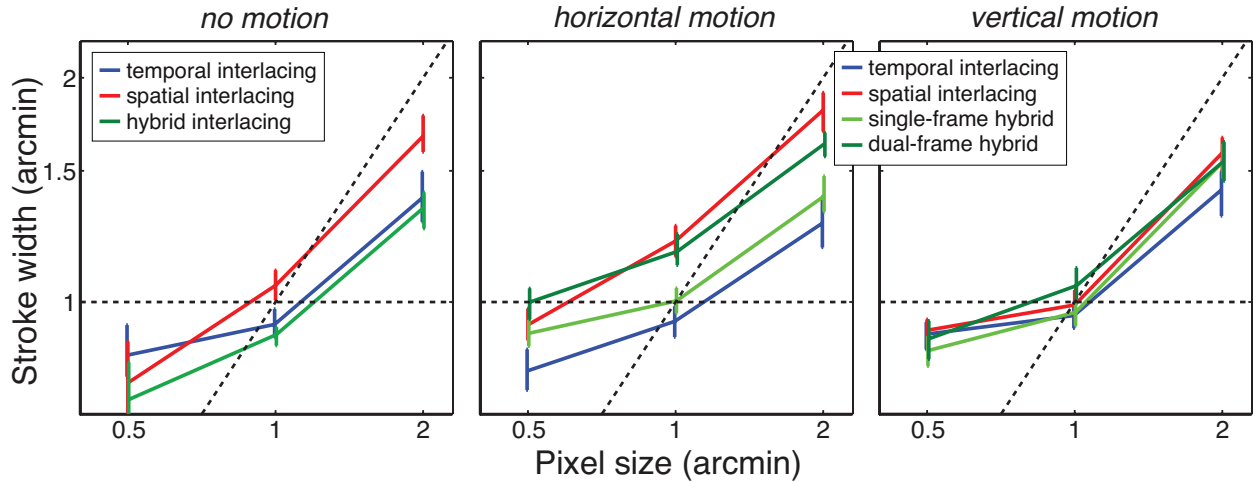


Figure 2.12: Effect of motion on spatial resolution. Data are averaged across subjects. The left, middle, and right panels show the data for no motion, horizontal motion, and vertical motion, respectively. Different colors represent the data from different protocols. Error bars represent 95% confidence intervals. The horizontal dashed line represents the expected resolution thresholds in the eye-limited regime and the diagonal dashed line the expected thresholds in the display-limited regime.

dashed lines represent the expected resolution thresholds for eye-limited and display-limited conditions, respectively. As you can see, thresholds increased as the simulated pixels became larger (i.e., as the simulated viewing distance became shorter) following the eye-limited and display-limited predictions fairly well. Importantly, resolution differed across protocols. Clearly, spatial interlacing had poorer effective resolution than temporal or hybrid interlacing in the display-limited regime (i.e., where the pixels were 1-2 arcmin). With smaller pixel sizes in the eye-limited regime (0.5 arcmin), the four protocols had very similar effective resolutions.

We next examined the influence of motion on effective spatial resolution. Figure 2.12 shows the data, averaged across observers, for stationary, horizontally moving, and vertically moving stimuli. With no motion, resolution with spatial interlacing was poorer than with the temporal-interlacing or the hybrid protocols at the shorter simulated viewing distances. With motion present, resolution with the spatial-interlacing and the dual-frame hybrid protocols was quite dependent on the direction of motion. When it was horizontal, resolution with those two protocols was worse than resolution with the temporal-interlacing and single-frame hybrid protocols. When motion was vertical, however, resolution with the spatial-interlacing and dual-frame hybrid protocols improved significantly. The improvement with vertical motion and the lack of improvement with horizontal motion both make sense. The problem with the spatial-interlacing and dual-frame hybrid protocols is that potentially useful data

are not shown to a given eye in every presentation. By moving the stimulus vertically, all parts of the letter can be presented to each eye over time, so performance improves. When the stimulus moves horizontally, the missing data are not presented at any time, so performance does not improve. The use of motion to create higher effective resolution has been examined extensively in computer graphics Didyk, Eisemann, Ritschel, Myszkowski, and Seidel, 2010.

We conclude that the single-frame version of the hybrid protocol has significantly better spatial resolution than the spatial-interlacing protocol. Indeed, the effective resolution of the proposed protocol is on par with temporal interlacing.

## 2.5 Experiment 4: Depth distortion

### Theory

In temporal interlacing S3D displays, the right- and left-eye views are presented alternately even though content is typically captured simultaneously. This means that the second eye sees an image that lags the correct timestamp. When there is movement in the scene, the visual system can interpret these temporal lags as disparity and the depth in the scene will be distorted Burr and Ross, 1979; Hoffman et al., 2011; Read and Cumming, 2005. This type of depth distortion should not occur with spatial-interlaced and hybrid displays because they present content simultaneously to the two eyes, yielding no ambiguity about which image in the right eye to match with a given image in the left eye.

Depth distortion as can occur in spatial interlacing displays should also be minimal in the hybrid technique. The proposed hybrid technique alternates the delivery of even and odd rows to the two eyes, so there is no consistent stimulus to drive vertical vergence to an unintended value. Thus, depth distortions due to the vertical offsets in spatial-interlacing displays should not occur with this technique. We did not test this possibility because Hakala et al. Hakala et al., 2015, in press have already shown that this type of distortion occurs with spatial interlacing and there is no reason to believe that it should occur with the hybrid protocols proposed here.

### Methods and apparatus

The same subjects and apparatus were used as in Experiment 1. The frame rate was 100Hz and the resolution of each display was  $1200 \times 900$ . We measured the perceived depth of moving objects in all four display protocols. We did this by presenting two sets of bright  $1^\circ$  squares moving horizontally in opposite directions against a dark background (Figure 2.13). Speed varied from  $-20$  to  $20^\circ/\text{sec}$ . The vertical edges of the squares were blurred slightly to reduce the salience of motion artifacts; this made the task easier to perform. Subjects were not instructed about fixation. After each trial they indicated whether the top or bottom squares appeared closer (2-alternative, forced-choice task). Based on the response, spatial disparity was added to the stimulus for the next trial according to a 1-down/1-up staircase

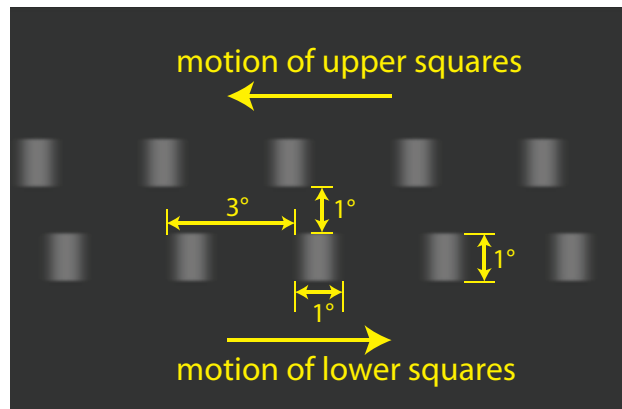


Figure 2.13: The stimulus used to measure depth distortion. Two groups of squares moved horizontally in opposite directions. On a given trial, the upper or lower group may have appeared closer than the other.

procedure. The goal was to find the disparity that had to be added to the moving stimulus in order to eliminate the depth distortion (i.e., make the top and bottom squares appear to be at the same depth). We call this added disparity the nulling disparity; it is a quantitative measure of the direction and size of the depth distortion.

The experiment consisted of about 320 trials per subject: 4 display protocols  $\times$  4 speeds  $\times$  20 trials. About 1 hour was required for each subject to complete the experiment. We determined psychometric functions using the method we described earlier.

## Results

The results are plotted in Figure 2.14. The nulling disparity—the spatial disparity required to eliminate depth distortion—is plotted as a function of object speed for the four protocols. As expected, large distortions of perceived depth occurred with temporal interlacing. Also as expected, the magnitude of the distortion was proportional to speed (Equation 1.1). The other protocols—spatial interlacing, dual-frame hybrid, and single-frame hybrid—yielded no depth distortion. We conclude that the hybrid techniques are immune to the depth distortions that plague temporal interlacing because the techniques present images simultaneously to the two eyes thereby allowing accurate disparity estimates.

## 2.6 Discussion

We found that the single-frame hybrid protocol maintains the benefits of both temporal and spatial interlacing, while eliminating the drawbacks. Specifically, motion appearance and flicker were significantly better than with temporal interlacing, depth distortion was

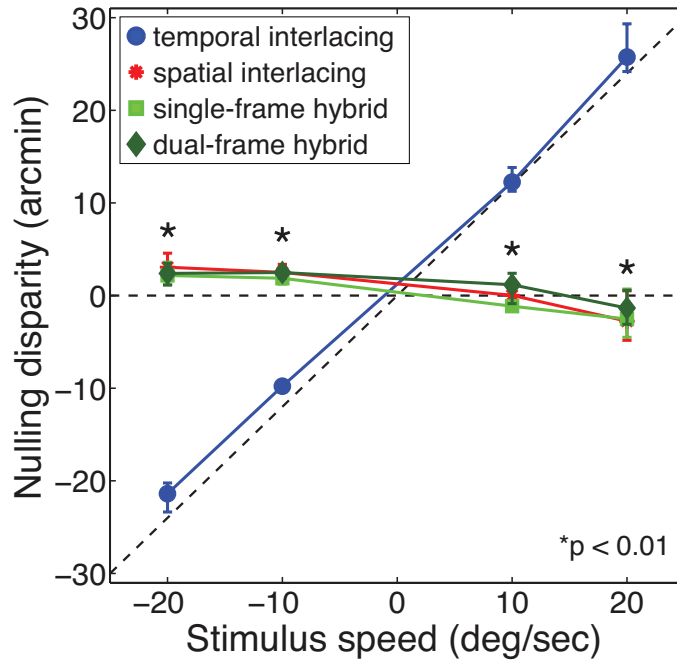


Figure 2.14: Depth distortions for different protocols. The nulling spatial disparity is plotted as a function of speed. Different colors represent the results from the different protocols. The data have been averaged across subjects. The diagonal dashed line represents the predictions of Equation 1.1 once multiplied by two because there were always two distortions in opposite directions, one for the rightward-moving stimulus group and one for the leftward-moving group. Error bars denote 95% confidence intervals. Asterisks indicate speeds at which the spatial-interlacing and hybrid protocols yielded significantly less distortion than the temporal-interlacing protocol (paired t-tests,  $p < 0.01$ ).

eliminated, and spatial resolution was better than with spatial interlacing. Thus, spatiotemporal interlacing is an attractive solution for presenting stereoscopic content with minimal temporal and spatial artifacts. We next discuss the underlying causes of the effects we observe with the different protocols and how one might implement the single-frame hybrid technique.

## Sampling, display, and viewing pipelines for different protocols

We have sufficient understanding of spatial and temporal filtering in the human visual system to make rigorous predictions about how different protocols, frame rates, duty cycles, and pixel sizes ought to affect flicker visibility, motion artifacts, and effective spatial resolution on a display. To this end, we modeled the pipeline from stimulus to display to viewing for the four protocols: temporal interlacing, spatial interlacing, and the two hybrid

techniques.

The display of video content involves three dimensions (two in space and one in time), but we show the analysis for two dimensions only (one in space and one in time) for ease of visualization. Typically image data  $i(x, t)$  are anti-aliased before being sent to the display, so we anti-aliased by convolving with a cubic interpolation function,  $a(x, t)$ . We then simulated how intensity varies over space and time when the image data are presented on a digital display. We sampled the anti-aliased image data with a comb function representing the spatiotemporal sampling of the display, where the samples are separated spatially by  $x_0$  (pixel spacing) and temporally by  $t_0$  (display frame time). The displayed intensities have finite spatial and temporal extent, which we represent with a spatiotemporal aperture function  $p(x, t)$ . The double asterisk represents two-dimensional convolution. In this example, the pixel fill factor is assumed to be 1 (meaning that the pixel width is equal to the inter-pixel separation), but the fill factor could have other values.

$$\left[ [i(x, t) ** a(x, t)] s(x, t) \right] ** p(x, y) \quad (2.1)$$

$$\left[ [i(x, t) ** a(x, t)] \text{comb}\left(\frac{x}{x_0}, \frac{t}{t_0}\right) \right] ** \text{rect}\left(\frac{x}{x_0}, \frac{t}{t_0}\right) \quad (2.2)$$

where  $\text{rect}$  is a scaled rectangle function with widths  $x_0$  in space and  $t_0$  in time, and  $x_0$  and  $t_0$  also represent the spatial and temporal separations of samples in the comb function. In the Fourier domain, the second equation becomes:

$$\left[ [I(f_x, f_t) A(f_x, f_t)] ** \text{comb}(x_0 f_x, t_0 f_t) \right] \text{sinc}(x_0 f_x, t_0 f_t) \quad (2.3)$$

where  $f_x$  and  $f_t$  are spatial and temporal frequency, respectively, and the  $\text{sinc}$  function has zeros at  $f_x = 1/x_0, 2/x_0, \text{etc.}$  and at  $f_t = 1/t_0, 2/t_0, \text{etc.}$  In the hybrid protocols, the sampling function has a phase shift in  $x$  at the alternation rate. In the single-frame hybrid protocol, there is also a phase shift in time. With these phase shifts, there are different spatiotemporal sampling functions. For the single-frame hybrid protocol, it is:

$$\text{comb}\left(\frac{x}{2x_0}, \frac{t}{2t_0}\right) \text{ for odd pixel rows} \quad (2.4)$$

$$\text{comb}\left(\frac{x + x_0}{2x_0}, \frac{t + t_0}{2t_0}\right) \text{ for even pixel rows} \quad (2.5)$$

and they alternate at the alternation rate. For the dual-frame hybrid protocol, the sampling function is:

$$\text{comb}\left(\frac{x}{2x_0}, \frac{t}{2t_0}\right) \text{ for odd pixel rows} \quad (2.6)$$

$$\text{comb}\left(\frac{x + x_0}{2x_0}, \frac{t}{2t_0}\right) \text{ for even pixel rows} \quad (2.7)$$

Because one set of pixel rows is presented with a delay of one display frame, we need separate spatiotemporal aperture functions for the odd and even pixel rows:

$$p\left(\frac{x}{x_0}, \frac{t}{t_0}\right) \text{ for odd pixel rows} \quad (2.8)$$

$$p\left(\frac{x}{x_0}, \frac{t - t_0}{t_0}\right) \text{ for even pixel rows} \quad (2.9)$$

In the simulations shown here, we assumed an illumination time equal to the display frame time and a pixel width and height equal to the pixel separation for the aperture functions. Other values could of course be assumed. We convolve the anti-aliased input with each of these sampling functions separately and then sum them to obtain an amplitude spectrum associated with each protocol. In the simulations shown here, we also assumed a display frame rate of 60Hz and a pixel size of 1 arcmin (because that corresponds to the recommendation for HD-TV and UHD-TV). Object speed was  $1.08^\circ/\text{sec}$  (65 pixels/sec). Other values could of course be assumed. We computed the amplitude spectra for one eye's image only because flicker and motion artifacts are determined primarily by monocular processing Hoffman et al., 2011 and effective spatial resolution is also determined primarily by monocular processing Kim and Banks, 2012.

The sequence of computations for a stimulus moving vertically across the screen at constant speed is shown in Figure 2.15. (We replace  $x$  with  $y$  in the figure to remind the reader that the motion is vertical.) The first row is a space-time plot of the stimulus and convolution with the anti-aliasing kernel. The output of the convolution is the second row and is the data sent to the screen to be displayed. The third row shows the sampling functions associated with the four protocols: from left to right, temporal interlacing, spatial interlacing, dual-frame hybrid, and single-frame hybrid. The fourth row represents the outputs of the sampling in the four protocols. Those outputs are convolved with the spatiotemporal function displayed in the fifth row to produce the space-time sequences of finite pixels at finite time intervals shown in the sixth row. Those sequences are subjected to Fourier transformation and the resulting spatiotemporal amplitude spectra are shown in the bottom row.

Figure 2.16 provides larger versions of the amplitude spectra for the four protocols. The spectra consist of a filtered version of the original signal (diagonal line through the origin) as well as spatiotemporal aliases. When aliases near the temporal-frequency axis are visible, viewers see flicker. When aliases in other locations in frequency space are visible, viewers typically see judder. The human visual system is sensitive to only a small range of the spatiotemporal frequencies generated by digital displays. The sensitivity range is quantified by the spatiotemporal contrast sensitivity function (also called the window of visibility Watson et al., 1986). The window of visibility is represented by the orange diamonds in each panel of Figure 2.16; the diamond shape is a reasonable approximation to the actual sensitivity function Kelly, 1979. When the aliases fall within the visible range, flicker and motion artifacts should be visible. When they fall outside the visible range, the stimulus

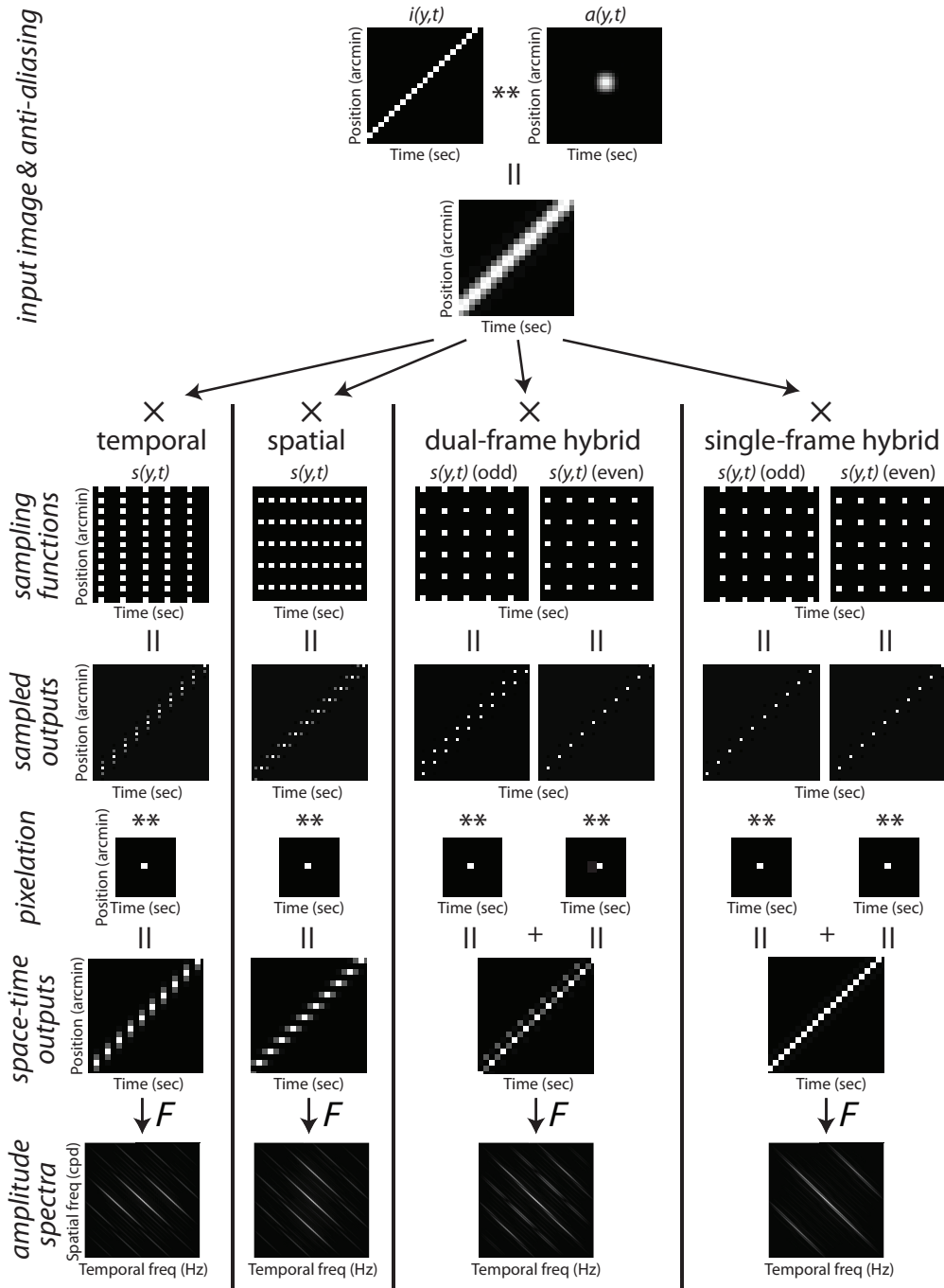


Figure 2.15: The presentation of a moving stimulus on the four protocols in space-time and in the Fourier domain. See text for details.

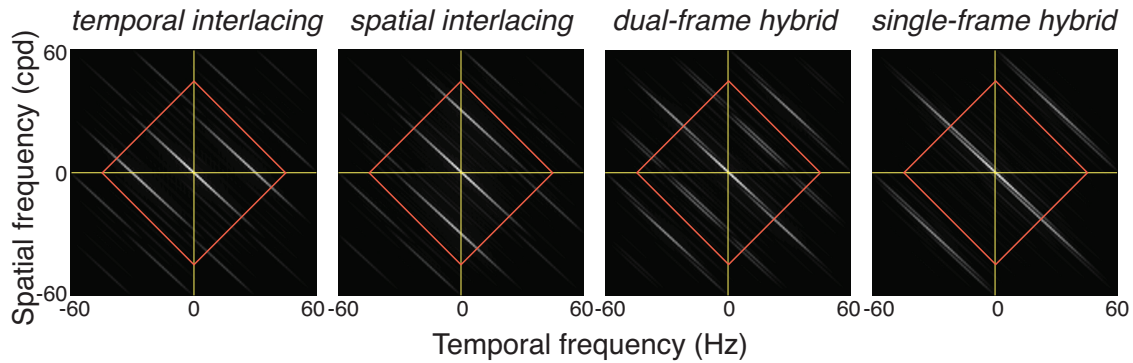


Figure 2.16: Amplitude spectra for a stimulus moving vertically in the four interlacing protocols. These panels are based on magnified versions of the bottom row in Figure 2.15. The spectra contain a filtered version of the original signal (diagonal line intersecting the origin) as well as aliases due to sampling. The orange diamonds represent the window of visibility, the range of spatial and temporal frequencies that are visible to a typical human observers. Aliases within the window of visibility can cause visible artifacts. In the case of temporal interlacing, aliases have large amplitudes in the temporal frequency direction, indicating the possibility of temporal artifacts, (e.g., flicker and motion artifacts). In spatial interlacing, aliases have large amplitudes in the spatial frequency direction, indicating a possible loss of spatial resolution. The single-frame hybrid has no replicates within the window of visibility, suggesting that aliases will not be visible.

should appear unflickering and its motion should appear smooth. The advantages of hybrid interlacing are readily apparent in the amplitude spectra. In particular, the single-frame hybrid should be relatively immune to flicker and motion artifacts because the aliases occur at higher frequencies than with the other protocols. This of course is what we observed experimentally.

This analysis of the pipeline also helps one understand the determinants of effective spatial resolution with different protocols and display parameters. In Figure 2.17 we present a hypothetical spatiotemporal stimulus with a low-pass amplitude spectrum; such a spectrum (represented by concentric circles at  $(0,0)$ ) is characteristic of most natural images. If the stimulus is presented on a non-interlacing display (first panel), aliases appear at every 60Hz in temporal frequency and 60cpd in spatial frequency. Temporal interlacing (second panel) loses half the frames for each eye’s view, resulting in additional aliases at temporal frequencies of -30 and 30Hz. Spatial interlacing (third panel) drops half the pixel rows in each eye’s view, yielding additional aliases at spatial frequencies of -30 and 30cpd. The single-frame hybrid protocol (fourth panel) produces aliases that are located diagonally in frequency space. This is because the sampling function in the space-time domain is a set of impulse functions positioned diagonally. Thus, the aliases created by hybrid interlacing are farther (by a factor of  $\sqrt{2}$ ) from the origin. The orange diamond in each panel represents the window of



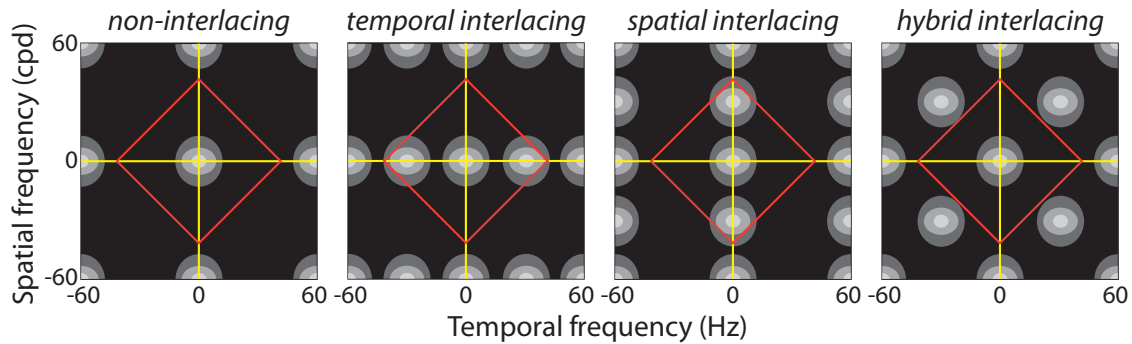


Figure 2.17: Amplitude spectra for different interlacing protocols. The display frame rate is 60Hz and pixel size is 1arcmin. The orange diamonds represent the window of visibility. The leftmost panel shows the signal presented on a non-interlacing display. The central pattern is the original signal, and its aliases repeat with a period of 60Hz in temporal frequency and 60cpd in spatial frequency. When temporal interlacing is used, the aliases occur at lower temporal frequencies (multiples of 30Hz). Similarly, when spatial interlacing is used, aliases occur at a lower spatial frequencies (multiples of 30cpd). Hybrid interlacing produces aliases at multiples of 30Hz in temporal frequency and 30cpd in spatial frequency, but they are located diagonally in frequency space. As a consequence, the aliases in hybrid interlacing are farther from the window of visibility, making them much less visible. In this cartoon, we omitted the effect of pixilation for simplicity.

visibility. Because of its diamond shape the aliases in the single-frame hybrid protocol are even less likely to be visible than in temporal and spatial interlacing. Thus, higher spatial frequencies can be seen by the viewer without intrusion by the aliases created in spatial interlacing. The prediction that the hybrid protocol should have higher spatial resolution than spatial interlacing was, of course, confirmed by our experimental measurements.

## Implementation of spatiotemporal interlacing

There are at least two ways to implement the hybrid protocol in a stereoscopic display. The first requires that the viewer have active eyewear that alternates left- and right-eye views; the second involves active polarization switching at the display and thereby allows one to use passive eyewear.

The first implementation is schematized in the left panel of Figure 2.18. The display sends light through a linear polarizing stage (yellow), which then transmits to a patterned quarterwave plate (gray). The quarterwave plate yields circular polarization that is clockwise in half the elements and counter-clockwise in the other half. The viewer wears active eyewear that alternates between the modes: one in which the clockwise elements are seen by the left eye and the counter-clockwise elements by the right eye (time 1) and one in which the

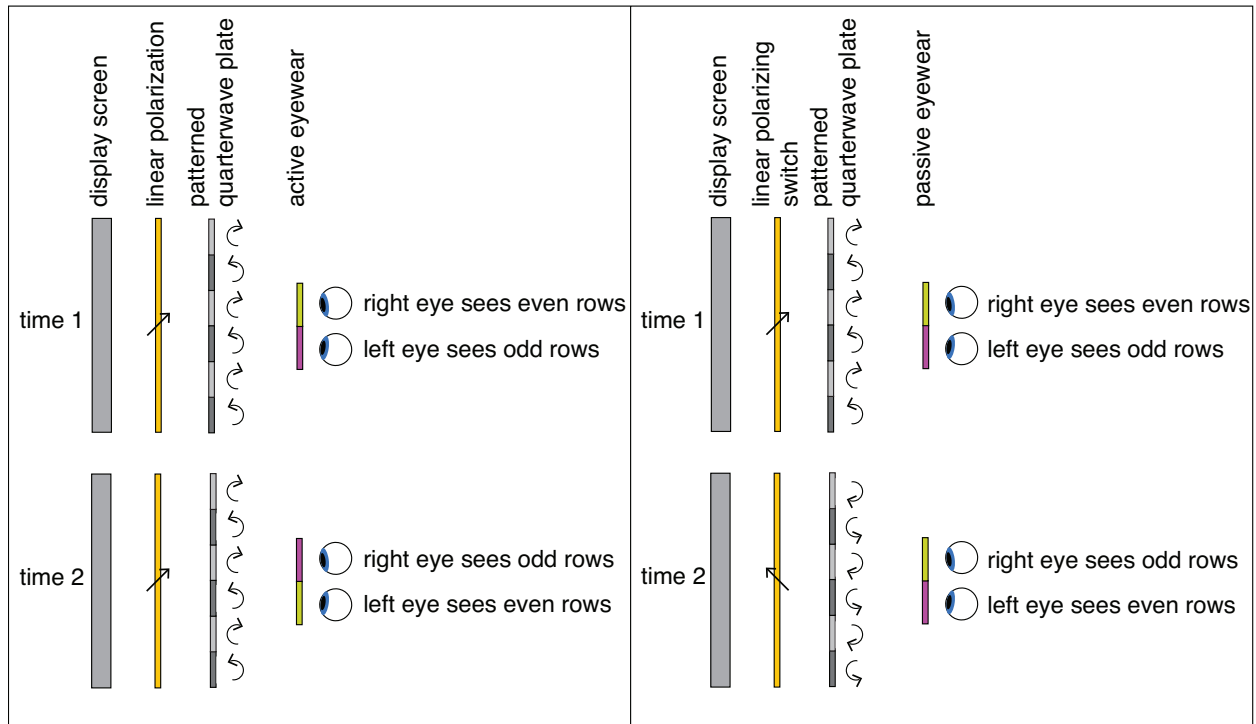


Figure 2.18: Possible implementation of the hybrid technique. The display could use a film-patterned retarder to first linearly polarize the light emitted from the display, similar to how a spatially interlaced television works. This light would then be sent through a wave plate that alternates quickly between  $1/4$  wave and  $3/4$  wave retardation, resulting in light that is circularly polarized either clockwise or counter-clockwise. The viewer would then wear passive polarized glasses to filter out the appropriate view at each moment in time.

clockwise elements are seen by the right eye and the counter-clockwise elements by the left eye (time 2).

It is undesirable in many applications to have active eyewear, so we designed an implementation that uses passive eyewear. This implementation is schematized in the right panel of Figure 2.18. The display sends light through a linear polarizing stage that switches between polarization angles of  $+45^\circ$  and  $-45^\circ$ . When the linear stage is at  $+45^\circ$ , the patterned quarterwave plate produces clockwise polarization in the odd rows and counter-clockwise in the even rows. When the linear stage is at  $-45^\circ$ , the quarterwave plate produces clockwise polarization in the even rows and counter-clockwise in the odd rows. The passive eyewear transmits clockwise polarization to the left eye and counter-clockwise to the right eye.

Both implementations would yield spatiotemporal interlacing as we have simulated in the experimental work presented here. The design of the quarterwave plate (e.g., row by row, checkerboard, etc.) determines the spatial pattern of the alternating blocks on the display.

## 2.7 Conclusion

In this chapter, I proposed a novel S3D presentation technique using spatiotemporal interlacing. Psychophysical experiments demonstrate that spatiotemporal hybrid interlacing maintains the better properties of both spatial and temporal interlacing. The hybrid technique has better spatial properties than spatial interlacing and better temporal properties than temporal interlacing. I developed a computational model that illustrates how different protocols ought to affect flicker, motion artifacts, and spatial resolution. The results from the model are consistent with the experimental results. I also provided a description of how this display might be implemented using currently available technology. This display technique should provide a better viewing experience than existing methods.

## Chapter 3

# Perceptual artifacts on a 240Hz OLED display

In this chapter I present research that investigated motion artifacts, flicker, and depth distortion on a 240Hz temporally interlaced S3D OLED display. The high frame rate of this display affords us some flexibility in how we deliver content to the two eyes, including the option of a dual-viewer mode that temporally interleaves 4 views, or 2 stereoscopic pairs. In a series of psychophysical experiments, we measured the visibility of artifacts on the OLED display using temporal interlacing and a 60Hz S3D LCD using spatial interlacing. We determined the relative contributions of the frame rate of the content, update rate of the display, duty cycle, and number of flashes. We found that short duty cycles and low flash numbers reduce the visibility of motion artifacts, while long duty cycles and high flash numbers reduce flicker visibility. We also tested several dual-viewer driving modes to determine the optimal mode for minimizing different kinds of artifacts. This work has been published in the *Journal of the Society for Information Display* Hoffman et al., 2015, in press; Johnson, Kim, Hoffman, and Banks, 2014; Johnson et al., 2015, in press.

### 3.1 Introduction

In stereoscopic 3D (S3D) displays, the method used to send left- and right-eye images to the appropriate eye can influence the visibility of artifacts. Temporally interlaced displays present left- and right-eye images alternately in time. Such interlacing has a maximum duty cycle of 0.5 because each eye only receives an image at most half of the time. We investigated duty cycles of 0.5 and less using an OLED display. To investigate a duty cycle of 1.0, we employed a spatially interlaced display. Spatially interlaced displays use a film-patterned retarder to present the left-eye image on even (or odd) rows and the right-eye image on odd (or even) rows. In this method, the two eyes are stimulated simultaneously, so one can generate a duty cycle of nearly 1.0. Thus, when tracking an object, motion blur should be more visible on a spatially interlaced display than on a temporally interlaced display.

Unlike LCDs, which have response times on the order of 4-9msec Elze and Tanner, 2012, OLED displays have a temporal response of less than 300usec in typical cases because they are limited only by the driving electronics Elze, Taylor, and Bex, 2013. OLED displays can thus be driven at high frame rates. A particular 240Hz OLED display prototype is capable of showing 240 unique frames per second and thus supports faster-than-normal capture rates and could thereby greatly reduce motion artifacts Hoffman et al., 2015, in press. The high frame rate also enables a dual-viewer S3D mode in which the four views needed for two viewers to see a left- and right-image pair are temporally interlaced on a single display. Two possible driving modes are  $L_A R_A L_B R_B$  and  $L_A L_B R_A R_B$ , where  $L_A$  and  $R_A$  are the left- and right-eye views for viewer A, and  $L_B$  and  $R_B$  are the left- and right-eye views for viewer B. We will refer to these protocols as LRXX and LXXR, respectively. The delay between the left- and right-eye views, or the interocular delay, is different in these two driving modes; LRXX has an interocular delay of 1/240sec while LXXR has an interocular delay of 1/120sec. Techniques have been proposed to predict and measure motion blur using digital measurement devices Feng, Pan, and Daly, 2008; Someya and Sugiura, 2007; A. B. Watson, 2010; O. J. Watson, 2013, and to create industry standards for the measurement of motion artifacts Miseli, 2006, but to the best of our knowledge there is no metric that can accurately predict the severity of multiple types of motion artifacts. We used a series of psychophysical experiments to measure the visibility of motion artifacts. Many of the effects we observed are consistent with an analysis of spatiotemporal signals in the frequency domain Adelson and Bergen, 1985; Hoffman et al., 2011; Watson et al., 1986.

## 3.2 Experiment 1: Motion artifacts

### Methods

To present S3D images, we used a prototype Samsung 240Hz OLED display that employs temporal interlacing and a commercially available LCD display (LG 47LM4700) that employs spatial interlacing. The diagonal lengths of the active areas of the OLED and LCD displays were 55 (1.40m) and 47 (1.19m) respectively. Viewing distance was 3.18 times picture height such that one pixel subtended 1arcmin, or 2.18 meters for the OLED display and 1.86 meters for the LCD display. Five subjects took part in the experiments. All had normal or corrected-to-normal vision. They wore the appropriate stereoscopic glasses for each display. On the temporally interlaced OLED, active shutter glasses were used, operating in one of two custom modes: left-right-left-right or left-left-right-right. On the spatially interlaced LCD, passive polarized glasses were used. The measurements were done both with stationary fixation and with tracking eye movements. For the LCD and OLED displays, we tested a range of capture and presentation protocols. In all, there were 18 conditions (nine presentation protocols with two eye-movement instructions). Figure 3.1 shows all the driving modes we tested. We presented 40 trials for each condition and speed, using the method of constant stimuli. For each type of stimulus, we asked the observer to report whether he or she perceived any

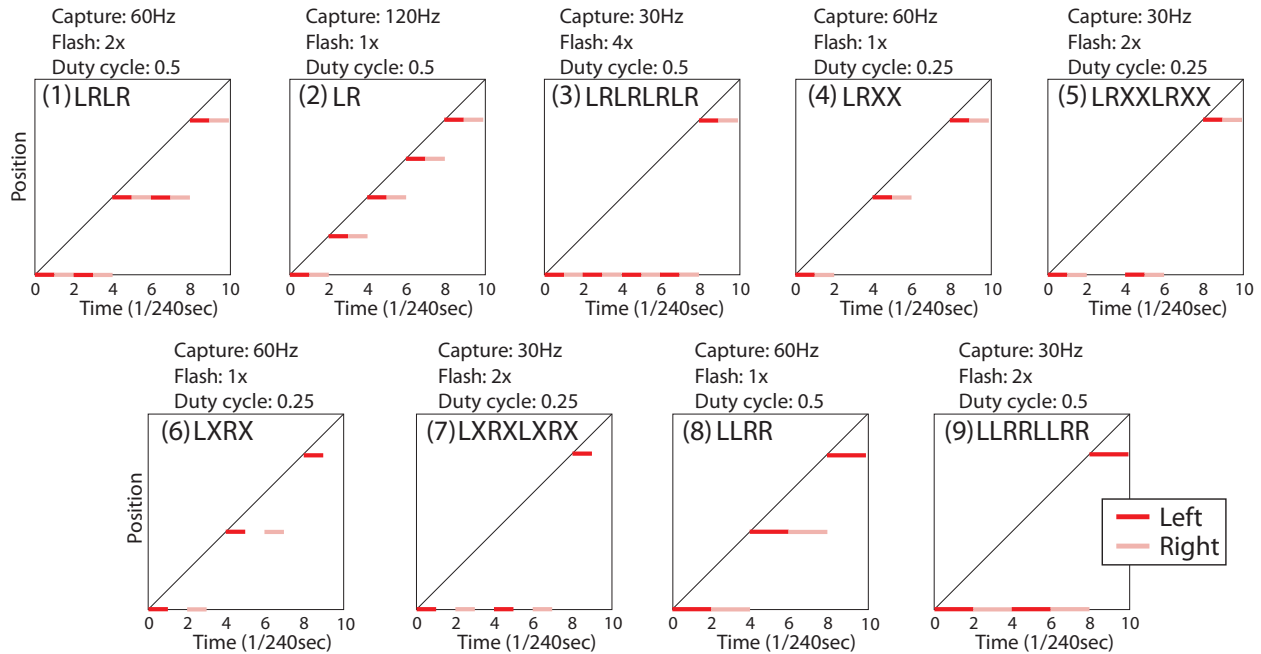


Figure 3.1: Driving modes presented on the 240Hz display with associated capture rate, flash number, and duty cycle. The gray diagonal line represents smooth continuous motion, and the horizontal red and tan lines represent left- and right-eye views, respectively.

motion artifacts, and we calculated the fraction of the 40 trials in which motion artifacts were reported. We fitted a cumulative Gaussian to the psychometric data using a maximum-likelihood criterion Fründ et al., 2011; Wichmann and Hill, 2001a, 2001b and extracted the object speed at which observers perceived motion artifacts half the time.

Figure 3.2 depicts the moving stimuli and fixation targets. In the tracking condition, the fixation target was initially off to one side, so the upcoming eye movement had to cross screen center. In the stationary condition, the fixation target was at screen center. The stimulus—a group of white squares moving horizontally at a constant speed—was visible for 1sec. Following the presentation, subjects reported whether or not they saw motion artifacts in the moving squares. Subjects were directed to respond regardless of the type of motion artifact perceived (i.e., blur, edge banding, or judder). It was often hard to articulate which type of motion artifact was present because they all can be present at once. Thus we focused on visibility of any motion artifact, rather than differentiating the types of artifacts.

## Results

Figure 3.3 shows the effect of capture rate on artifact visibility on the OLED display for all five observers. Each panel plots for a different subject the object speed at which

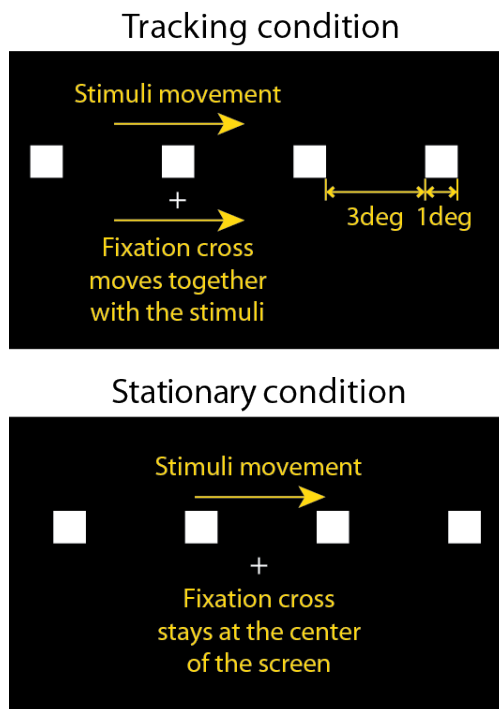


Figure 3.2: Stimulus and fixation target in the tracking and stationary conditions. A trial consisted of three parts: initial fixation, stimulus motion, and response collection. In the tracking condition, the fixation target moved with the same velocity as the squares across the center of the display. In the stationary condition, the fixation target remained stationary.

artifacts were visible as a function of capture rate. Thresholds generally increased with capture rate up to the maximum rate of 120Hz. There is noticeable inter-subject variability in the stationary condition at high capture rates, but observers were fairly consistent in their own artifact ratings. For simplicity in subsequent plots, thresholds are averaged across subjects and the error bars represent the standard deviation for the observers.

One of our core experimental questions concerned the difference between the presentation rate and capture rate. We examined how (single, double, and quadruple) flashing affects the visibility of motion artifacts in order to evaluate the assertion that strobed presentation can improve the quality of perceived motion. Figure 3.4 shows data averaged across subjects, for the stationary and tracking conditions, and demonstrates the relationship between the number of flashes and motion artifacts. There was a clear effect of capture rate on artifact visibility in both the stationary and tracking cases. At the lowest capture rate of 30Hz, we tested the double- and quadruple-flash protocols only because single flash had unacceptable flicker. There was no significant difference between double and quadruple flash with 30Hz capture. At 60Hz capture, we could only test single- and double-flash protocols. There was no significant benefit of single flash over double flash in the stationary condition, but in the

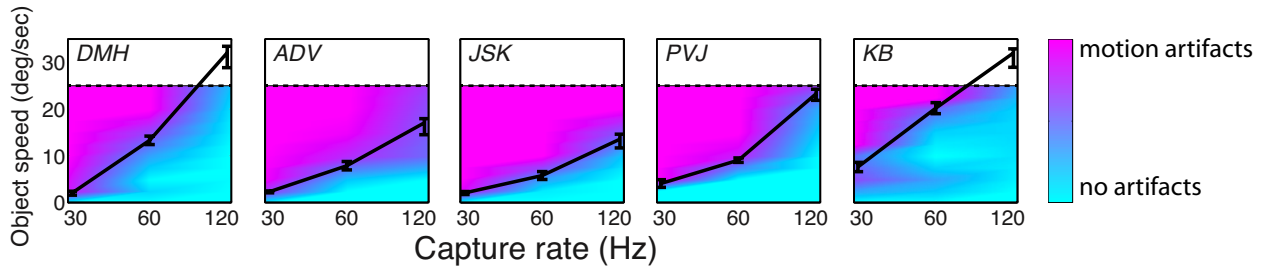


Figure 3.3: Variation between observers and effect of capture rate on motion artifacts for the stationary fixation condition. Object speed at which artifacts were reported on half the trials is plotted as a function of capture rate. Thus, greater ordinate values indicate fewer motion artifacts. Each panel shows the data from one subject. Interocular delay was  $1/240$ sec. Presentation was single flash except for the 30Hz capture rate, which was double flash. Protocols correspond to #2, 4, and 5 in Figure 3.1. Error bars represent 95% confidence intervals. The fastest speed tested was  $25^\circ/\text{sec}$  (dashed line), so any thresholds above that value are an extrapolation. DMH, ADV, JSK, and PVJ were authors; KB was not.

tracking condition, motion was significantly smoother with single flash than double flash (paired t-test,  $p < 0.01$ ). These results—no difference during stationary fixation and large differences during tracking—are consistent with the predictions of the retinal-position model in Figure 1.3. In other words, artifacts in the stationary condition were more likely caused by judder, while artifacts in the tracking condition were more likely caused by motion blur or edge banding.

We can also carry out a similar analysis to assess impact of the duty cycle of the presentation. Figure 3.5 shows the results for duty cycles of  $\sim 0.25$ ,  $\sim 0.5$ , and  $\sim 1.0$  with a capture rate of 60Hz. A spatially interlaced display was used for the duty cycle of 1.0. In the stationary condition, duty cycle had no significant effect on motion artifacts. In the tracking condition, there was a clear effect of duty cycle: The  $\sim 1.0$  duty cycle caused motion artifacts at approximately half the speed of the  $\sim 0.5$  duty cycle presentation. The shortest duty cycle of  $\sim 0.25$  supported the fastest motion without artifacts. This effect in the tracking condition was due to the increase in motion blur with larger duty cycles.

We measured the effect of interocular delay on the visibility of motion artifacts by comparing the LRXX and LXX protocols. Figure 3.6 shows the results. There was no systematic effect of interocular delay on motion artifact visibility. This finding is consistent with the experiments by Hoffman and colleagues who concluded that the visibility of motion artifacts is determined by monocular signals Hoffman et al., 2011. Interocular delays have an influence, however, on other effects such as depth distortion and flicker Hoffman et al., 2011; Hoffman et al., 2015, in press.



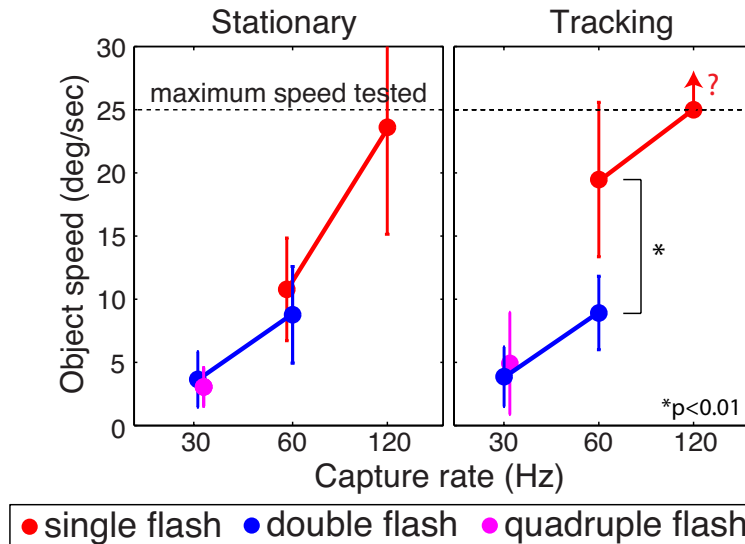


Figure 3.4: Effect of flash number on motion artifacts. The data have been averaged across the five subjects. The object speed at which artifacts occurred half the time is plotted as a function of capture rate. The left and right panels correspond to stationary and tracking conditions, respectively. The dashed horizontal line represents the maximum speed tested. Thresholds that lie on that line indicate that no motion artifacts were observed at the fastest speed tested; extrapolating the exact threshold in those cases would be unjustified. Error bars indicate one standard deviation from the mean. The protocols tested correspond to protocols #1, 2, 3, 8, and 9 in Figure 3.1. All protocols have duty cycle 0.5. The comparison marked with a bracket indicates a significant difference as evidenced by a paired t-test ( $p < 0.01$ ).

### 3.3 Experiment 2: Flicker

#### Methods

To measure flicker thresholds on the OLED display, we used the same setup as in Experiment 1. Four subjects with normal or corrected-to-normal vision took part in the experiment. The visual system is somewhat more sensitive to flicker in the peripheral than in the central visual field Tyler, 1987. For this reason, we presented stimuli in the periphery in this experiment in order to obtain a worst-case estimate of flicker visibility. Subjects viewed the display from half the normal viewing distance and fixated on a cross 20 arcmin from the top of the screen. A solid gray rectangle was presented for 1 sec in the lower center of the screen, subtending an angle on the retina of  $20^\circ$  (horizontal) by  $13.3^\circ$  (vertical). In retinal coordinates, the stimulus was located between  $22.7^\circ$  and  $35.7^\circ$  in the peripheral visual field, which is approximately where Tyler Tyler, 1987 found the highest flicker fusion frequencies.

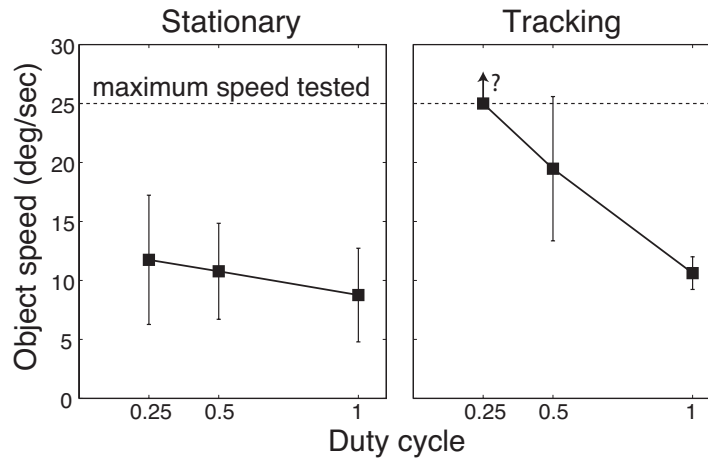


Figure 3.5: Effect of duty cycle on motion artifacts. The left panel corresponds to the stationary condition and the right panel to tracking. The object speed at which artifacts were reported on half the trials is plotted as a function of duty cycle. The capture rate was 60Hz and presentation was single flash. Duty cycles of 0.25 and 0.5 were presented on the temporally interlaced display, and the duty cycle of 1.0 on the spatially interlaced display. We excluded one subject’s data in the right panel because we could not fit the psychometric function to some conditions. Error bars represent one standard deviation. Some error bars are too small to see in this plot. The temporal interlacing protocols tested correspond to protocols #6 and 8 in Figure 3.1.

Subjects indicated whether the rectangle appeared to flicker. We presented stimuli with different luminance values using a staircase to determine the point at which subjects perceived flicker half the time. All single- and dual-view modes were tested.

## Results

We measured the visibility of flicker for the different driving modes. Figure 3.7 shows flicker thresholds as a function of display protocol. Thresholds represent the luminance above which a large bright object in the peripheral vision appeared to flicker. There was a small decrease in flicker visibility when the left- and right-eye images were 180 degrees out of phase (LXX) as opposed to 90 degrees out of phase (LRXX). A long duty cycle (LLRR) decreased flicker visibility further. A double flash protocol (LRLR) had no visible flicker whatsoever, even when the display was at maximum screen brightness. In this case, the 120Hz fundamental frame rate per eye is well above the critical flicker frequency of the visual system, even in peripheral areas of the visual field. The spatially interlaced display had no visible flicker (data not shown).

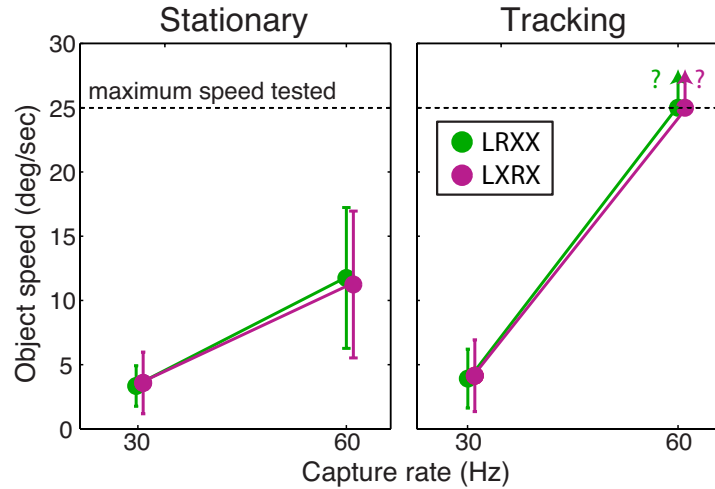


Figure 3.6: Effect of interocular delay on motion artifacts. The slowest object speed at which artifacts were reported on half the trials is plotted as a function of capture rate, for stationary (left) and tracking (right) conditions. Green and magenta circles correspond to LRXX and LXRX protocols, respectively (protocols #4-7 in Figure 3.1). Error bars represent one standard deviation.

### 3.4 Experiment 3: Depth distortion

A temporal delay between the left- and right-eye views can be interpreted as disparity, as previously described Hoffman et al., 2011; Read and Cumming, 2005. We investigated the driving modes possible on the 240Hz OLED to determine the optimal way to deliver content to the two eyes to minimize depth distortion. The LRXX mode has an interocular delay of  $1/240$ sec while LXRX has an interocular delay of  $1/120$ sec. We would therefore predict that depth distortion would be more severe in the LXRX mode than the LRXX mode.

#### Methods

We measured depth distortion on the same OLED and LCD displays from Experiment 1. Four subjects with normal or corrected-to-normal vision took part in the experiment. The stimulus consisted of two groups of vertical bars drifting in opposite horizontal directions, similar to the stimulus used in Figure 2.13 to measure depth distortion in the hybrid display protocol. Because the distortion is proportional to the horizontal velocity of the target, one group of bars appeared to be closer while the other group appeared to be farther than the display surface. The task was to choose the closer group using a keyboard. It was a 2-alternative forced choice task: the subjects had to decide even when they were not sure about their percepts. A 1-up, 1-down staircase procedure found the nulling disparity necessary to make the top and bottom appear at the same depth. We tested two capture

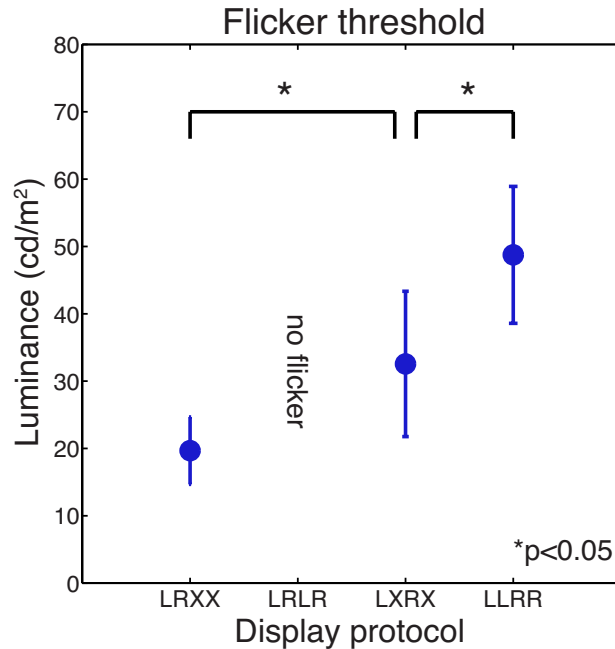


Figure 3.7: Flicker thresholds averaged across the subjects. Thresholds signify the luminance value above which flicker is perceived for a large bright object in the peripheral visual field. Comparisons marked with the brackets at the top indicate significant differences as evidenced by a paired t-test ( $p < 0.05$ ). Error bars represent one standard deviation.

conditions: simultaneous capture and alternating capture. In the simultaneous capture condition, left- and right-eye views are captured simultaneously and presented alternatingly. In the alternating capture condition, left- and right-eye views were captured at the appropriate timestamp for when each eye is presented. We fitted a cumulative Gaussian to the psychometric data using a maximum-likelihood criterion Fründ et al., 2011; Wichmann and Hill, 2001a, 2001b.

## Results

We found the disparity necessary to eliminate depth distortion for different capture conditions and interocular delays. The top two panels of Figure 3.8 show the effect of interocular delay on depth distortion on the OLED display, with data pooled across all subjects. When capture was simultaneous, a larger interocular delay resulted in a greater amount of depth distortion, roughly in line with predictions. Predictions are based on the speed of the stimulus and interocular delay, as in Equation 1.1. When capture was alternating, there was little distortion of depth. There appeared to be a global shift upwards of  $\sim 1$ -2 arcmin, possibly due to the vertical horopter; subjects are more likely to perceive the top squares as behind

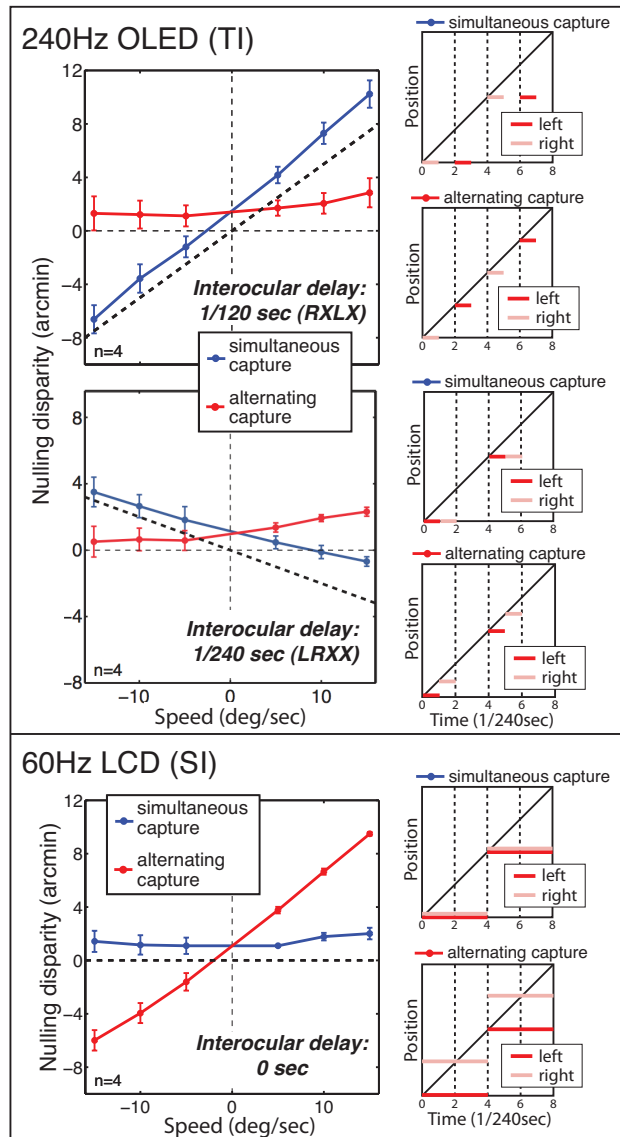


Figure 3.8: Effect of interocular delay on depth distortion. Data have been pooled across subjects. Interocular delay in the top panel is 1/120sec, 1/240sec in the middle panel, and zero in the bottom panel (60Hz FPR spatially interlaced display). Note that the magnitude is reversed in the middle panel because of the ordering of left/right presentation (RXLX vs. LRXX). Blue lines represent results from the simultaneous capture condition, while red lines represent the alternating capture condition. The panels on the right illustrate the protocols being tested. Error bars represent 95% confidence intervals. Depth distortion is markedly worse with simultaneous capture (when presentation is alternating), and also worse with a larger interocular delay.

the bottom squares in depth because the vertical horopter is tilted away from fronto-parallel Cooper, Burge, and Banks, 2011; Schreiber, Hillis, Filippini, Schor, and Banks, 2008. The bottom panel of Figure 3.8 shows the data from the FPR LCD display. When presentation was alternating, as in the OLED display (top panel), alternating capture minimized depth distortion. When presentation was simultaneous, as in the LCD display (bottom panel), simultaneous capture minimized depth distortion.

## 3.5 Discussion

### Motion and flicker

We have shown that higher capture rates yield fewer motion artifacts, but that capture rate is not the only predictor of such artifacts. We also showed that a longer duty cycle yields more motion blur if the viewer is tracking a moving object, but fewer artifacts than in the stationary case, even for a duty cycle near 1.0. Generally, subjects were more sensitive to motion artifacts in the stationary condition than the tracking condition. In typical cases, viewers will most likely track salient objects in the scene and therefore be substantially less likely to attend to objects outside of fixation that may suffer from judder.

To explain why judder is worse during stationary fixation compared to tracking, consider the signal in Fourier space when the viewer tracks a moving object. When we plot the retinal position of a moving object as a function of time, as in Figure 3.9, the object moves across the retina in the stationary condition, but remains still in the tracking condition. If the slope of the continuously moving object is  $s$ , then the slope of the signal and replicates in Fourier space is  $-1/s$ . The slope influences how much of the replicate energy falls within the window of visibility in the stationary condition. As the speed of the object increases, the replicates tip further and intrude deeper into the window of visibility, causing more severe judder. In the tracking condition, replicates are sheared such that they become vertical (assuming perfect tracking), which has the same effect as slowing down the stimulus. This reduces the extent to which replicates fall within the window of visibility and therefore reduces judder.

Motion blur, on the other hand, is due to the sample-and-hold property of OLED and LCD displays. The retinal-position hypothesis provides one explanation for why a longer duty cycle increases motion blur, but an analysis of signals in Fourier space can provide insight into why this happens. A sample-and-hold protocol with a duty cycle of 0.5 is used to present the motion schematized in Figure 3.9. We can think of a sample-and-hold protocol as stroboscopic sampling convolved with a rect function. In frequency space, that has the effect of multiplying by a *sinc* function. The sharpness of an object is determined by high-spatial-frequency information near a temporal frequency of 0. In the stationary case, the *sinc* function is oriented vertically with a peak-to-trough distance of  $1/(d\Delta t)$  in the horizontal (temporal frequency) direction, where  $d$  is the duty cycle and  $\Delta t$  is the capture period. The *sinc* envelope has no effect on high spatial frequencies when the temporal frequency is low, so duty cycle does not create motion blur in the stationary case. In the tracking

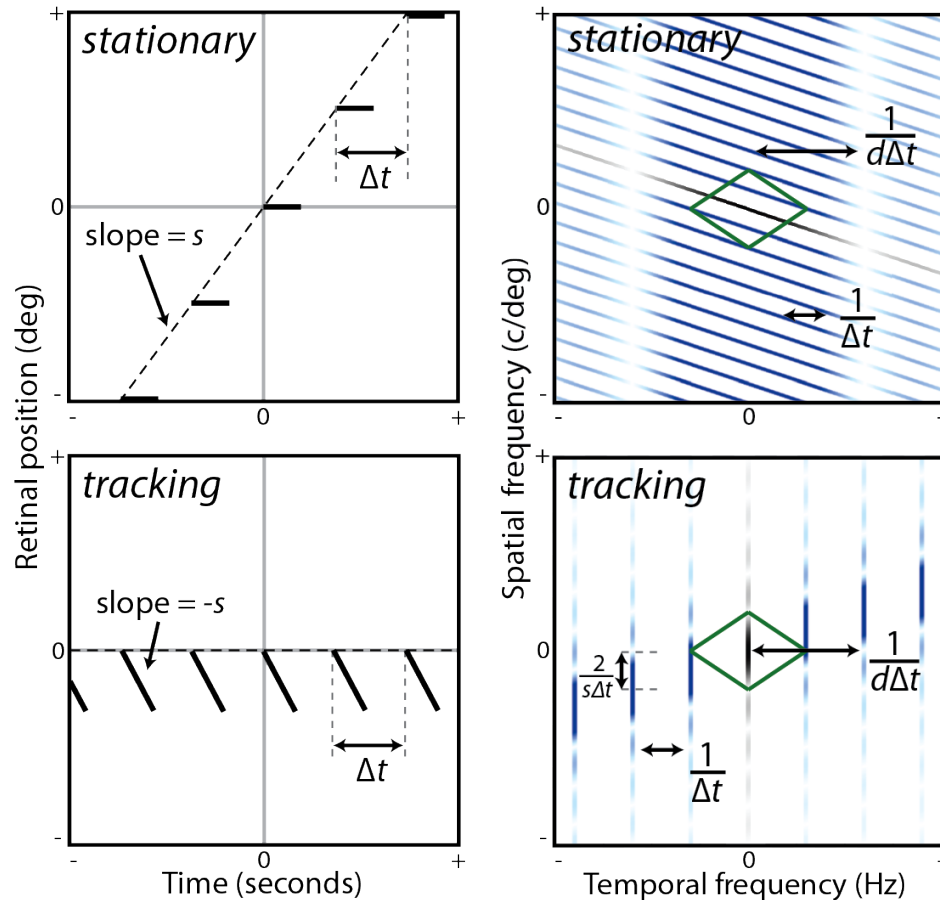


Figure 3.9: Effect of eye movements on the perception of judder and blur. The top panels correspond to the stationary fixation condition and the bottom panels to the tracking condition. The black lines in the left panels show the retinal position over time of a smoothly moving object presented using a sample-and-hold display. The right panels show the resulting amplitude spectra of the continuous signal (black) and replicates (blue). The diamond represents the window of visibility. In the stationary condition, replicates are located within the window of visibility, causing judder. In the tracking condition, replicates remain outside the window of visibility, but high-spatial-frequency information in the signal has been lost due to sample-and-hold presentation.

case, however, the *sinc* function is sheared vertically, which has the effect of attenuating high spatial frequencies at a temporal frequency of 0, causing motion blur. Furthermore, the spread of the *sinc* function in frequency space is a function of the duty cycle as well as speed of the object; the peak-to-trough distance is  $1/(ds\Delta t)$  in the vertical (spatial frequency) direction, and remains  $1/(d\Delta t)$  in the horizontal direction. In Figure 3.9 the duty cycle is 0.5, which would produce a vertical spread of  $2/(s\Delta t)$ . With a lower duty cycle of 0.25, this distance would be  $4/(s\Delta t)$ , spreading the *sinc* further in the vertical direction, reducing the attenuation of high spatial frequencies within the window of visibility. This would make blur due to motion less apparent. The width of motion blur can also be expressed in units of retinal distance rather than frequency units. In this case, the width of the blur,  $b$ , can be expressed using the following equation:

$$b = \frac{f - 1 + d}{f} s\Delta t \quad (3.1)$$

where  $f$  is the number of flashes,  $d$  is duty cycle,  $s$  is object speed, and  $\Delta t$  is capture period. Note that for multiple flashes, the blur width is confounded by the fact that other artifacts such as edge banding may be visible, but this equation provides an upper limit for the retinal blur that can occur.

It is important to consider the normal range of object speeds in typical content. A study of Japanese households and broadcast content estimated the viewing conditions and content that people typically experience in their homes Fujine, Kikuchi, and Sugino, 2007. Based on knowledge of how far people sit from their TVs and the motion in broadcast content, they found speeds of less than  $10^\circ/\text{sec}$  in 40% of scenes,  $10\text{-}20^\circ/\text{sec}$  in 30% of scenes, and  $20\text{-}40^\circ/\text{sec}$  in 30% of scenes. This finding, combined with our result that motion artifacts are generally visible in the range of  $10\text{-}20^\circ/\text{sec}$  when the capture rate is 60Hz, suggests that a capture rate of 60Hz is inadequate to create smooth motion in typical scenes. Particularly when viewers are fixating on a static part of the scene, they are likely to experience significant artifacts.

Presentation rates of 60Hz per eye or higher are used in displays to avoid visible flicker on moderately bright displays. Sample-and-hold displays, including OLEDs and LCDs, do not have such a strict requirement because the long duty cycle has the effect of attenuating spatiotemporal aliases in the frequency domain. Regardless, these displays are traditionally driven at 60Hz per eye or higher to create reasonably smooth motion. However, temporal interlacing for S3D lowers the duty cycle and makes flicker an important consideration. Frame rates must therefore be higher than an equivalent non-stereoscopic display. Our results demonstrate that 60Hz presentation is inadequate to completely eliminate flicker in peripheral vision for any of the dual-viewer modes. However, the 240Hz OLED display has a high enough frame rate to afford some flexibility in how stereoscopic 60Hz content is presented in single-viewer mode. If eliminating flicker is a priority, then content could be presented with double flash (LRLR). If eliminating motion artifacts is a priority, content could be presented with single flash and the lowest possible duty cycle of  $\sim 0.25$  (LXR or



LRXX) to reduce blur. In this case, flicker could be noticeable in certain types of content, particularly when there are large areas of high luminance.

Multiple-flash protocols, while helpful for minimizing flicker, can cause artifacts of their own. In digital 3D cinema, the popular RealD format presents 24Hz content using a triple-flash display protocol for a presentation rate of 72Hz. This triple-flash technique ensures that the presentation rate is high enough to avoid visible flicker. In S3D cinema, left- and right-eye views are interleaved temporally for a presentation rate of 72Hz per eye or 144Hz overall. This driving scheme produces obvious motion artifacts, predominantly edge banding. However, attempts to move to higher capture rates—such as in Peter Jackson’s *The Hobbit*, filmed (capture rate) at 48fps—have received mixed feedback. Many viewers complain of a so-called soap opera effect that causes content to feel less cinematic, like a made-for-TV movie Marks, 2012. To the best of our knowledge, this effect has not been rigorously characterized. An important consideration could also be the shutter function used to capture content. For a capture rate of 24Hz used in cinema, the shutter is kept open for a long time to increase motion blur, which makes motion appear smoother for content that would otherwise suffer from extreme judder Hoffman et al., 2011. Computer games are typically rendered without motion blur and thus have many sharp moving edges that are prone to judder. Reconsidering the shutter function for high capture rates could provide benefits. If the shutter function in the filming of *The Hobbit* was the same proportion of the frame capture period, its duration would have been half the duration of standard 24Hz capture, thereby decreasing motion blur and increasing judder.

These experiments have shown some large differences in how motion artifacts are perceived depending on eye movements, capture rate, and duty cycle. The dual-viewer modes supported by the 240Hz OLED display are effective at producing fewer motion artifacts than spatially interlaced displays largely due to differences in the duty cycle, even though they are slightly more susceptible to visible flicker than either of the single-viewer modes. It is also worth considering that the spatially interlaced display used in this study is an LCD, not an OLED. Compared to OLED displays, LCDs have slower, and asymmetric, rise and fall times and are therefore less temporally precise Cooper, Jiang, Vildavski, and Farrell, 2013. LCD response times can even exceed one frame Elze, 2010. This could result in greater amounts of motion blur due to image persistence on the screen, independent of the fact that the duty cycle in a spatially interlaced display is already greater than would be permitted in a temporally interlaced display. Some recent technologies have been introduced to speed up the liquid-crystal response time (e.g. dynamic capacitance compensation), but these techniques can often cause artifacts of their own Elze and Tanner, 2012. Heesch and colleagues showed that the temporal aperture, or the temporal extent of the pixel aperture, can be used to predict flicker, motion blur, and judder Heesch and Klompenhouwer, 2008. The work analyzed the effect of the temporal aperture on spatiotemporal aliases to show that short duty cycles reduce the appearance of blur but increase the visibility of flicker. This is consistent with our result.

The finding that dual-viewer strategy did not influence perceived motion artifacts confirms that the visibility of motion artifacts is primarily dictated by the monocular images;

i.e., there is little if any effect of the phase of stimulation between the two eyes Cavonius, 1979; Hoffman et al., 2011. This is not the case, however, with flicker visibility. The phase of stimulation between the two eyes appears to play a role in flicker perception; there was a slight benefit of LXX over LRXX. Previous studies have shown that flicker-fusion rates are higher when illuminated frames are presented in phase in the two eyes, compared to when they are presented 180 degrees out of phase Hoffman et al., 2011; Cavonius, 1979. It therefore makes sense that frames presented 90 degrees out of phase would cause a similar increase in flicker visibility.

### Depth distortion

It is also worth considering the fact that the temporal delay between left- and right-eye inputs often creates distortions in the perceived depth of moving objects because temporal delay is interpreted as spatial disparity. We confirmed previous work Read and Cumming, 2005 that had shown that a longer interocular delay causes more depth distortion. The LRXX driving mode (interocular delay 1/240s) therefore has at least one benefit over the LXX mode (interocular delay 1/120s). The method in which content is captured (i.e. alternating vs. simultaneous) also plays a crucial role. We verified that alternating capture eliminates depth distortion in temporally interlaced displays, while simultaneous capture eliminates depth distortion in spatially interlaced (simultaneous) displays Hoffman et al., 2011. Furthermore, any method that reduces interocular delay should reduce depth distortion. The 240Hz display allows for a shorter interocular delay using the LRXX mode than any possible driving mode in temporally interlaced displays running at 120Hz, the current standard for this technique.

## 3.6 Impact

This work assesses how a variety of display-related factors can influence the visibility of artifacts. The strongest factor influencing motion artifacts is the frame rate of the content depicted on the display. OLED technology offers rapid response times such that the bottleneck of the imaging system is no longer pixel response time. It is now possible to take advantage of multi-viewer temporal interlacing and new approaches to generate content at high frame rates. One such method to extend the benefits of high-frame-rate displays is the development of improved motion-compensated, frame-rate conversion routines Methods and systems for improving low resolution and low frame rate video, 2008 Systems and methods for a motion compensated picture rate converter, 2012. These routines use sophisticated computer-vision algorithms that track the movement of objects in a scene and interpolate between consecutive frames to fill in the missing frames. The calculation of high-quality interpolated frames can have a substantial impact on reducing artifacts for fast-moving objects.

The discussion of motion clarity has been clouded by the widespread adoption of LCD displays with LED backlights. LEDs can be used to strobe the LCD display faster than the

refresh rate, effectively creating a multi-flash driving mode intended to lower the sample-and-hold duty cycle of the display. Many display manufacturers report the LED backlight strobing frequency rather than the true refresh rate of the display, claiming to have refresh rates as high as 1440Hz, even though the true refresh rate of the displays is much lower, at 120 or 240Hz. Song and colleagues showed that motion blur is reduced on strobed-backlight LCDs compared to continuous-backlight LCDs due to the lower backlight duty cycle, but they did not provide a metric for the possible edge banding that could occur Song, Teunissen, Li, and Zhang, 2009. Some manufacturers offer a flicker-free mode for some of their OLED monitors, in which the signal is switched on and off twice or more within one frame, equivalent to multiple flash Cooper et al., 2013; Elze et al., 2013. Though flicker may be reduced in this case, our research shows the potential downside of multiple-flash techniques in that they can exacerbate banding artifacts. A 240Hz display with a backlight strobing at 1440Hz does not increase the capture rate of the content and is therefore unlikely to substantially improve the appearance of motion compared to a 240Hz display with a continuous-backlight LED. Samsung’s Clear Motion Rate, LG’s Motion Clarity Index, and Sony’s MotionFlow all report refresh rates significantly higher than the real refresh rate of the display.

Another method that has been proposed to reduce motion blur on sample-and-hold displays (both LCD and OLED) is black-data insertion Klompenhouwer, 2006; Song et al., 2009. By doubling the frame rate and inserting a blank frame after each frame, this effectively reduces the duty cycle from  $\sim 1.0$  to  $\sim 0.5$ . Shortening the duty cycle would be particularly easy for OLED displays because they have an immediate temporal response. Our research provides evidence that this driving mode should reduce the presence of motion artifacts. However, the display would be more susceptible to flicker, and the display would require a higher light output to negate the dimming effect of the black frames.

### 3.7 Conclusion

We examined a 240Hz OLED display and found that a low flash number and low duty cycle reduces artifact visibility under tracking conditions with flicker being slightly more visible. This finding, combined with a clear benefit of higher capture rate, provides evidence to support the move to higher frame rate in television as well as cinema, which can utilize a lower flash number if the content has a higher frame rate. Our results also emphasize the importance of developing content for high-frame-rate displays.

## Chapter 4

# The visibility of color breakup

Color breakup is an artifact seen on displays that present colors sequentially. When the eye tracks a moving object on such a display, different colors land on different places on the retina, and this gives rise to visible color fringes at the object's leading and trailing edges. Interestingly, color breakup is also observed when the eye is stationary and an object moves by. Using a novel psychophysical procedure, we measured breakup both when viewers tracked and did not track a moving object. Breakup was somewhat more visible in the tracking than in the non-tracking condition. The video frames contained three sub-frames, one each for red, green, and blue. We spatially offset the green and blue stimuli in the second and third sub-frames respectively to find the values that minimized breakup. In the tracking and non-tracking conditions, spatial offsets of  $\Delta x/3$  in the second sub-frame (where  $\Delta x$  is the displacement of the object in one frame) and  $2\Delta x/3$  in the third eliminated breakup. Thus, this method offers a way to minimize or even eliminate breakup whether the viewer is tracking or not. We suggest ways to implement the method with real video content. We also developed a color-breakup model based on spatio-temporal filtering in color-opponent pathways in early vision. We found close agreement between the model's predictions and the experimental results. The model can be used to predict breakup for a wide variety of conditions. This work has been published in the *Journal of Vision* Johnson, Kim, and Banks, 2014.

### 4.1 Introduction

Figure 4.1 depicts the individual sub-frames presented to the eyes in the RGB protocol (R in the first sub-frame, G in the second, and B in the third). The stimulus is a four-pixel-wide white bar moving at constant speed. When the viewer fixates a stationary point (non-tracking), the images of the moving object from the three sub-frames land on the same position on the retina, but at different times. When the viewer tracks the object (tracking), the three sub-frames present images at the same location on the display, but different locations on the retina.

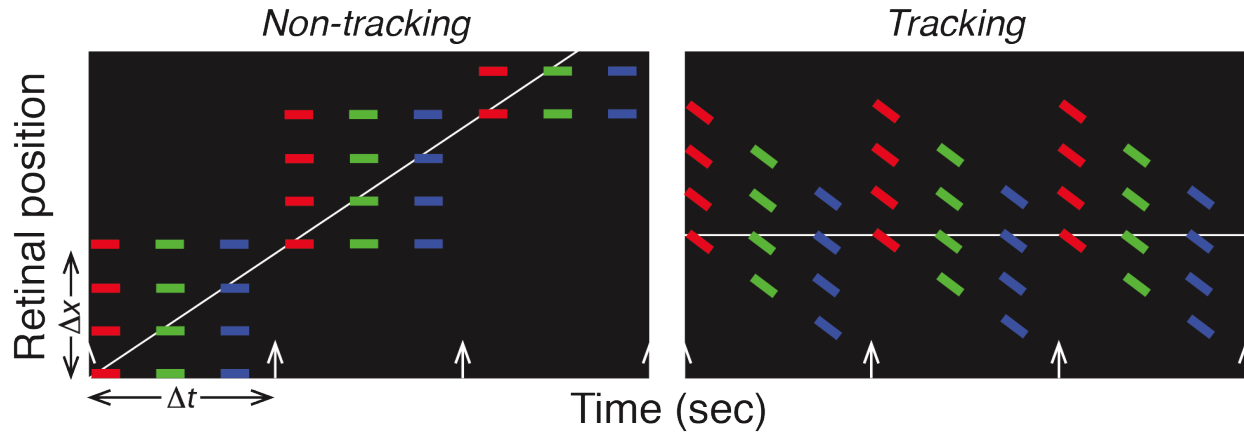


Figure 4.1: Space-time plot of the retinal images generated by a narrow white bar moving at constant speed. Position on the retina is plotted as a function of time. The object is four pixels wide and moving at three pixels per frame. The object's displacement on the screen is  $\Delta x$  for each frame of duration  $\Delta t$ . Left: The viewer fixates a stationary point while the object moves. The sub-frames within a frame fall on the same retinal position but at different times. Right: The viewer tracks the object. The red component spatially leads the green by  $\Delta x/3$  due to the eye movement during the first sub-frame. The green component also leads the blue component by  $\Delta x/3$ .

To understand the color-sequential stimulus, it is useful to consider the spatio-temporal differences between a bright moving stimulus with R, G, and B presented sequentially versus simultaneously: i.e., R, G, and B are presented together in each of three sub-frames. The sequential presentation requires stimulation in all three sub-frames, so we made the same assumption for simultaneous color presentation. Let the frame duration be  $\Delta t$  and the displacement of the stimulus on the screen  $\Delta x$ . Figure 4.2 shows the spatio-temporal modulation of red and green when the stimulus is composed of equiluminant R, G, and B. The left column plots the modulation when the viewer does not track the stimulus and the right column the modulation when the viewer tracks it. The upper row shows the modulation in R as a function of space and time and the lower row the modulation in G. (The panel for B would be very similar, just shifted one sub-frame.) In the figure, gray represents no difference in luminance between the sequential and simultaneous presentations. White represents space-time positions in which the luminance of R (or G) is greater in the sequential than in the simultaneous stimulus. Black represents space-time positions in which R (or G) is lower. As you can see, the differences with tracking are identical to those without tracking except for a vertical shear parallel to the position axis; the shear, of course, is due to the eye movement. Thus, from the physical differences in luminance—that is, from the differences without consideration of spatio-temporal filtering by the visual system—one cannot predict whether tracking or non-tracking will produce more apparent color breakup.

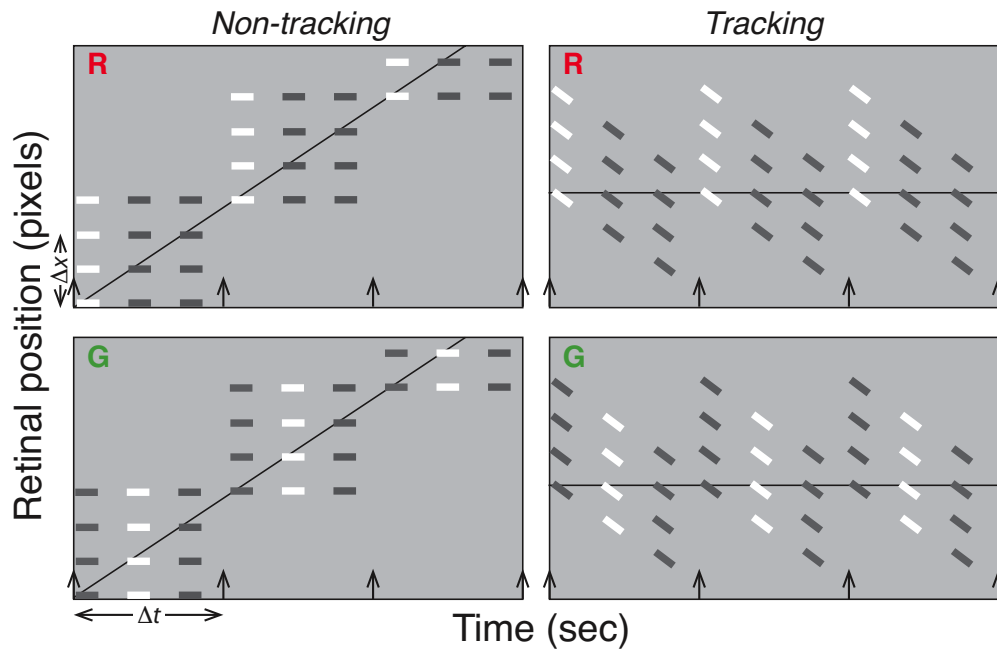


Figure 4.2: Space-time differences in color modulation between simultaneous and sequential color presentation. Each panel plots the variation in red or green stimulation as a function of time and retinal position when a white object moving at a speed of  $\Delta x/\Delta t$  is presented (two pixels per frame). The object is four pixels wide. Vertical arrows indicate frames, each consisting of three sub-frames. In the sequential presentation, the sub-frames are R, then G, and then B. In the simultaneous presentation, R, G, and B are presented in all three sub-frames. The left column shows the differences when the viewer does not track the object and the right column the differences when the viewer tracks. The upper and lower rows show the differences for R and G, respectively. Gray represents no difference. White represents cases in which R (or G) is more luminous in the sequential presentation. If the luminance of R (and G) is 1 integrated across a frame, then white represents  $+2/3$ . Black represents cases in which R (or G) is less luminous, and corresponds to  $-1/3$ .

With tracking, introducing a spatial offset of  $\Delta x/3$  in the image in the second sub-frame relative to the first and an offset of  $2\Delta x/3$  in the third sub-frame relative to the first will cause the sub-frames to be presented at the same location on the retina. Figure 4.3 shows differences in the modulation of red and green between sequential and simultaneous presentation for an offset of  $\Delta x/3$  in the G (second) sub-frame. The modulation differences with and without tracking are identical except for a vertical shear parallel to the position axis. In fact, the differences are identical to those without the spatial offset except for shifts parallel to the position axis due to the offset. The modulation differences in space-time, therefore, do not reveal whether spatial offsetting should reduce or eliminate color breakup nor whether offsetting should have a different effect with tracking than without tracking. This means that color breakup is not an attribute of the display or video content, but rather is created in visual processing. In particular, without low-pass filtering in early vision, color breakup would always occur when red, green, and blue are presented sequentially and breakup would be unaffected by spatial offsets of the sort illustrated in Figure 4.3.

## 4.2 Model of color breakup saliency

To understand the causes of color breakup and particularly ways to minimize its visibility, we need to incorporate spatio-temporal filtering in color-opponent channels in early vision (Figure 4.4). We constructed a model based on measured spatial and temporal properties of color-opponent processes in the human visual system. It is designed to predict the color-opponent response modulation that the stimuli in our experiments would create. We assume that the visibility of color breakup is monotonically related to the response modulation.

The stimulus was a bright achromatic rectangle on a dark background. The rectangle moved horizontally, so we considered only the horizontal dimension of space. We first defined the input stimulus in two-dimensional space-time coordinates as in Figure 4.1. The input is in retinal coordinates. We were interested mainly in the chromaticity of the percept, so we converted the RGB values of the input stimuli into values in opponent color space, which approximates the luminance and chromatic channels of the visual system Poirson and Wandell, 1993; Zhang and Wandell, 1997. First, we linearized the RGB values. Because the stimulus in our simulation was a white target on a black background, we set the linearized RGB to  $[1\ 1\ 1]$  in the target and  $[0\ 0\ 0]$  in the background. We then transformed the linearized RGB values into tristimulus values in the CIE XYZ color space. If the values conform to sRGB IEC61966-2.1, a standard color space, the transformation is Winkler, 2005:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412 & 0.358 & 0.180 \\ 0.213 & 0.715 & 0.072 \\ 0.019 & 0.119 & 0.950 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4.1)$$

The transformation from CIE XYZ color space to opponent color space is Zhang and Wandell, 1997:

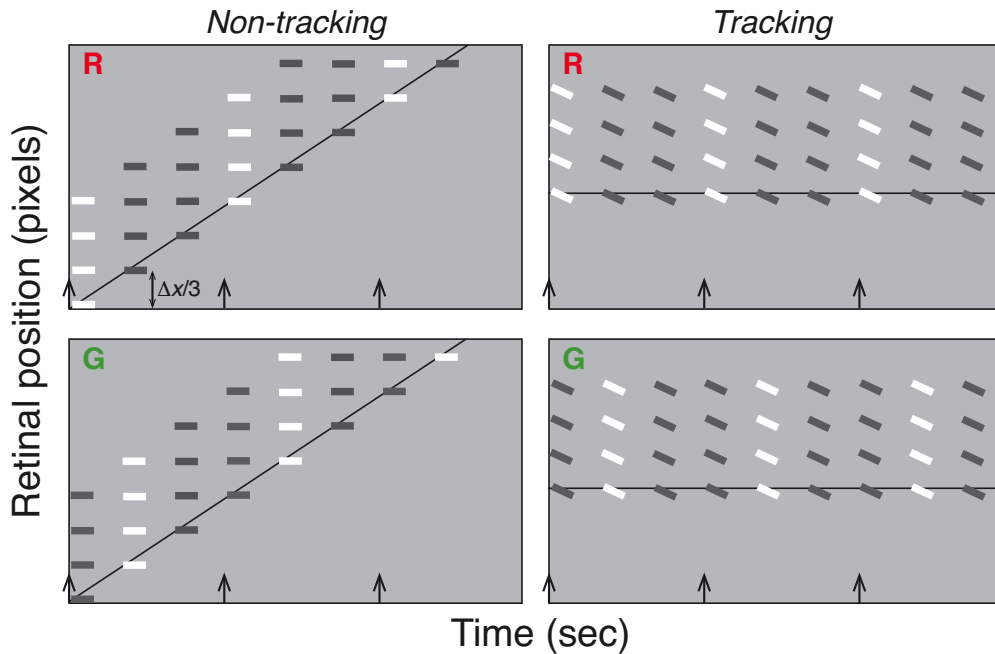
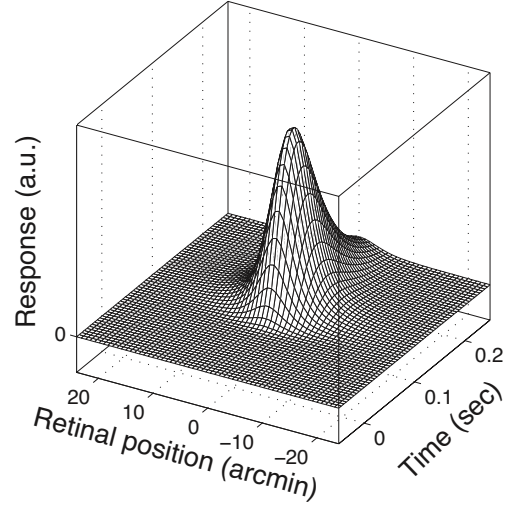
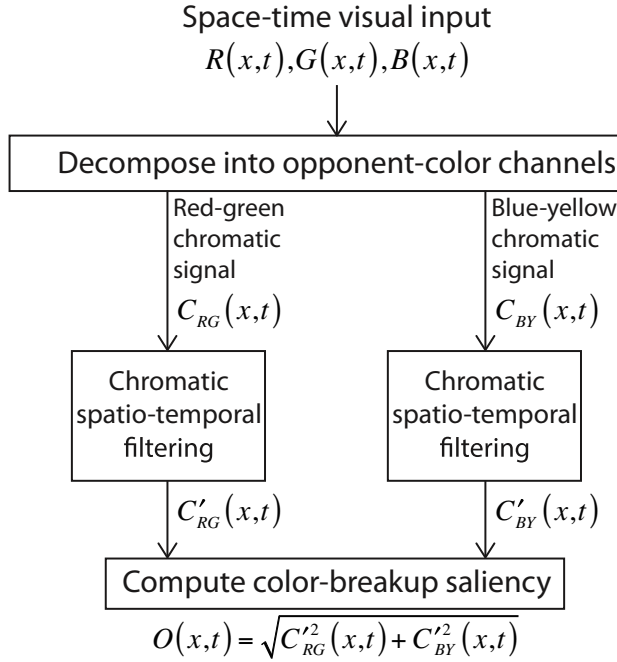


Figure 4.3: Space-time color modulation when a spatial offset of  $\Delta x/3$  is applied to the G sub-frame. Each panel plots the variation in red or green as a function of time and retinal position when a moving white object is presented. In the sequential presentation, the sub-frames are R, then G, and then B. In the simultaneous presentation, R, G, and B are presented in all three sub-frames. The left and right columns show the variation without and with tracking, respectively. The upper and lower rows show the variation for the R and G channels, respectively. Gray represents no difference. White represents cases in which R (or G) is more luminous in the sequential presentation. Black represents cases in which R (or G) is less luminous.

$$\begin{bmatrix} A \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.279 & 0.720 & 0.107 \\ -0.449 & 0.290 & 0.077 \\ 0.086 & 0.590 & 0.501 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (4.2)$$

As the displayed image changes over time, the luminance of pixels changes accordingly. Ideally, temporal changes would be instantaneous, but the display has its own temporal response. We incorporated the temporal properties of the display (ViewSonic G225f) by measuring its temporal impulse-response function (IRF). The function was very close to an ideal exponential decay with a time constant of 1.5msec. All simulation results included the display's IRF. (When we replaced the IRF with a delta function, the simulation results were nearly identical, which means that the display's IRF was short relative to the assumed human IRF.) The resulting values in opponent color space, passed through the display's IRF,





(A) Simulation of color-breakup saliency

(B) IRF for chromatic channel

Figure 4.4: Model of color breakup. The model estimates the visibility of color breakup based on spatio-temporal filtering in color-opponent channels in early vision. The left side of the figure depicts the model's components. The right side shows the impulse-response function (IRF) for the color-opponent channels used in the model.

were then convolved with the IRFs of the two chromatic channels. To calculate IRFs, we adopted the red-green channel's contrast sensitivity function (CSF) as described by Burbeck and Kelly Burbeck and Kelly, 1980 and Kelly Kelly, 1983 (Equation 4.3). The CSF is the sum of excitatory (E) and inhibitory (I) components; it is not separable in the space-time domain even though the excitatory and inhibitory components are separable.  $E_s$ ,  $E_t$ ,  $I_s$ , and  $I_t$  are defined respectively as the spatial response of the excitatory component, temporal response of the excitatory component, spatial response of the inhibitory component, and temporal response of the inhibitory component. The last equation under Equation 4.3 is used commonly in deriving the other equations in 4.3.  $f_{s1}$ ,  $f_{s2}$ ,  $f_{t1}$ , and  $f_{t2}$  are constants that depend on individual variances. We adopted values for  $f_{s1}$ ,  $f_{s2}$ ,  $f_{t1}$ , and  $f_{t2}$  measured by Burbeck and Kelly Burbeck and Kelly, 1980: 10cpd, 0.5cpd, 19Hz, and 1Hz.

$$CSF(f_x, f_t) = E(f_x, f_t) + I(f_x, f_t) = \frac{E_s(f_x)E_t(f_t)}{E_s(f_{x2})} + \frac{I_s(f_x)I_t(f_t)}{I_s(f_{x2})} \quad (4.3)$$

$$E_s(f_x) = \begin{cases} \frac{K(f_{x1}, f_{t2})}{K(f_{x1}, f_{t1})} K(f_x, f_{t1}) & \text{for } f_x \leq f_{x1} \\ K(f_x, f_{t2}) & \text{for } f_x > f_{x1} \end{cases} \quad (4.4)$$

$$E_t(f_x) = \begin{cases} \frac{K(f_{x2}, f_{t1})}{K(f_{x1}, f_{t1})} K(f_{x1}, f_t) & \text{for } f_t \leq f_{t1} \\ K(f_{x2}, f_t) & \text{for } f_t > f_{t1} \end{cases} \quad (4.5)$$

$$I_s(f_x) = E_s(f_x) - K(f_x, f_t) \quad (4.6)$$

$$I_t(f_x) = E_t(f_x) - K(f_x, f_t) \quad (4.7)$$

$$K(f_x, f_t) = 4\pi^2 f_x f_t \left[ 6.1 + 7.3 \left| \log_{10} \left( \frac{f_t}{3f_x} \right) \right|^3 \right] e^{-4\pi \frac{2f_x + f_t}{45.9}} \quad (4.8)$$

The CSF has no phase information, so we had to reconstruct phase, which we did by extending the assumption of a minimum-phase filter Kelly, 1969; Mason and Zimmerman, 1960; Stork and Falk, 1987 to the space-time domain. Such an assumption is reasonable for chromatic channels Burr and Morrone, 1993. We assume a complex transfer function,  $H$ , which is the Fourier transform of the IRF of the visual system,  $h$ .

$$\int \int h(x, t) e^{-2\pi i(f_x x + f_t t)} dx dt = H(f_x, f_t) = |H(f_x, f_t)| e^{i\theta(f_x, f_t)}. \quad (4.9)$$

Here, the modulus is the same as the CSF. Our goal was to estimate  $\theta(f_x, f_t)$  in order to reverse-engineer  $h$ . First, we computed the logarithm of the transfer function. The log of the modulus and the unknown phase become respectively the real and imaginary parts:

$$\log[H(f_x, f_t)] = \log[|H(f_x, f_t)|] + i\theta(f_x, f_t) \quad (4.10)$$

If these real and imaginary parts are a Hilbert transform pair, then  $h$  is real and causal Mason and Zimmerman, 1960. The Hilbert transform pair is not the only solution for causal and real  $h$ , but it is the solution that satisfies minimum phase. For a signal  $k$  whose independent variable is  $t$ , the Hilbert transform is

$$K_{Hi}(t) = \frac{-1}{\pi t} * k(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{k(t') dt'}{t' - t} \quad (4.11)$$

where the Cauchy principal value takes care of the integration around the singularity at  $t = t'$ . In our case, we want a causal signal in time such that

$$h(x, t) = 0 \text{ for } t < 0 \quad (4.12)$$

However, we do not need such a constraint in the space dimension. Thus we apply the Hilbert transform along the temporal frequency dimension only. The convolution and transforming integration that we used were:

$$\theta(f_x, f_t) = \left[ \frac{-\delta(f_x)}{\pi f_i} \right] * \log[|H(f_x, f_t)|] = \frac{1}{\pi} \int \int \frac{\delta(f_x - f'_x)}{f'_t - f_t} \log[|H(f'_x, f'_t)|] df'_x f'_t \quad (4.13)$$

where  $\delta$  is the Kronecker delta function. With the phase term estimated, we calculated the IRF  $h(x, t)$  by taking the inverse Fourier transform of  $H(f_x, f_t)$ . The calculated IRF was causal and real. We also used the same IRF for the yellow-blue channel because it is very similar to its red-green analog Mullen, 1985; Tong, Heeger, and van den Branden, 1999.

Figure 4.5 shows some of the modeling results. Each panel plots predicted modulation among color-opponent channels as a function of retinal position and time; when the value is zero there is no modulation, so breakup should not be seen. The left half of the figure shows the responses when the viewer's eyes are stationary and the right half the responses when the viewer tracks the stimulus. The panels from top to bottom show the responses when the second (green) sub-frame is displaced relative to the first (red) sub-frame by  $\Delta x/3$ , 0,  $\Delta x/3$ ,  $2\Delta x/3$ , and  $\Delta x$ , respectively, and the third (blue) sub-frame is displaced by twice those values. As expected from consideration of low-pass temporal filtering, a spatial offset of  $\Delta x/3$  eliminates color breakup in the model when the viewer tracks the object. Interestingly, the same offset minimizes predicted breakup when the eyes are stationary. The range of offsets for which breakup should not be visible is wider in the non-tracking case than in the tracking case.

As we said earlier, the cause of visible breakup during tracking when the offset is zero is obvious: different sub-frames fall on different positions on the retina and therefore appear spatially displaced. The cause of breakup with zero offset when the eyes are stationary is less obvious. When a frame is presented, three sub-frames land on one position on the retina, but at different times. When the next frame appears, the visual impression from the first sub-frame of the last frame has faded more than the impression from the second and third sub-frames. Therefore, a preceding frame affects the appearance of the first sub-frame the most and a succeeding frame affects the last sub-frame the most. At the trailing edge, a black frame follows a white frame leaving the visual impression from the last sub-frame, which is blue. At the leading edge, a black frame precedes a white frame, so the visual impression from the first sub-frame, which is red, is more salient than from the other sub-frames. As a result, one sees blue and red at the trailing and leading edges, respectively: the same ordering as observed when tracking the white object. A spatial offset of  $\Delta x/3$  for the second sub-frame and  $2\Delta x/3$  for the third causes the trailing and leading edges to oscillate mostly in hue. Sensitivity to rapid variations in hue is low, so the visibility of breakup is reduced or eliminated.

We predict that a spatial offset of  $\Delta x/3$  of the second sub-frame, and  $2\Delta x/3$  for the third sub-frame, can minimize color breakup regardless of whether the viewer is tracking a moving object or fixating a stationary point while an object moves past. We also predict that the range of offsets that eliminate or minimize color breakup will be wider when the eyes are stationary than when they are tracking. We conducted psychophysical experiments to evaluate these predictions.

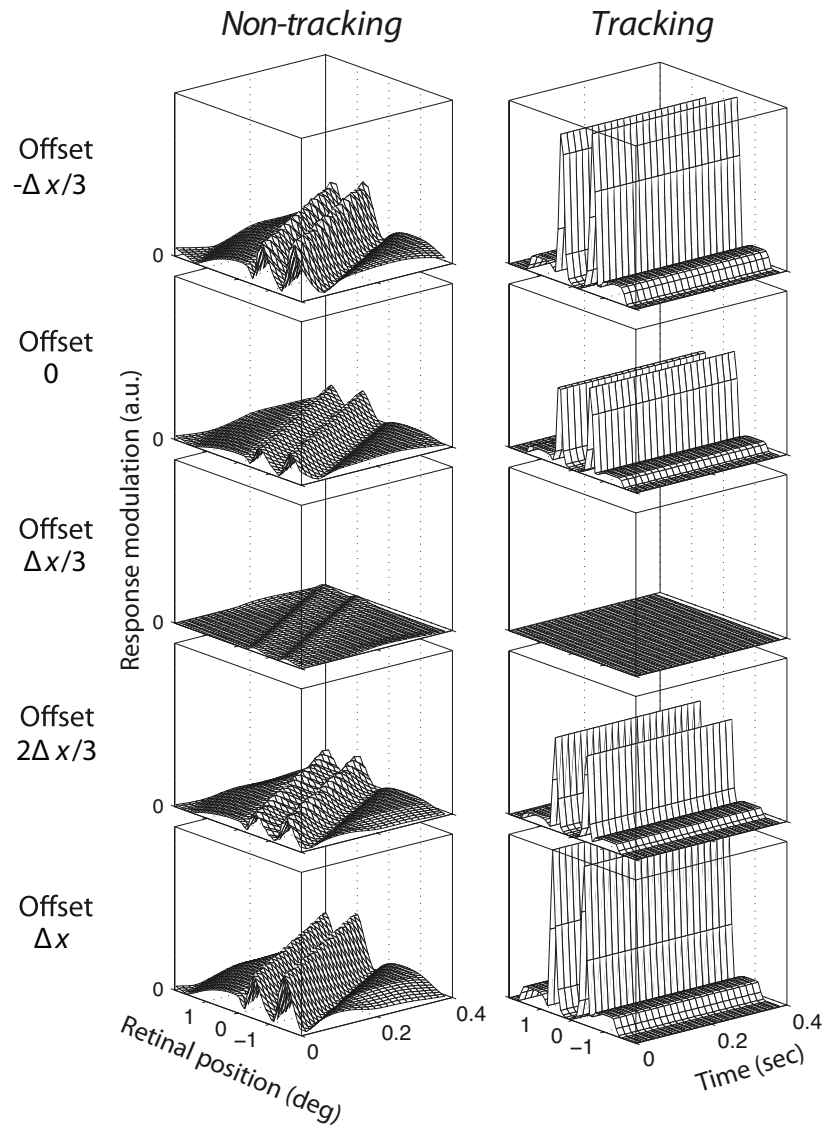


Figure 4.5: Model predictions. The panels plot the output of the model described in Figure 4.4 for various conditions. Each panel plots predicted response modulation among color-opponent channels as a function of retinal position and time. Zero represents no response modulation and therefore no predicted breakup. Greater values correspond to greater predicted modulation and therefore breakup. The left and right columns represent respectively the responses when the viewer's eyes are stationary and when the viewer tracks the stimulus. From top to bottom, the spatial offsets of the second sub-frame are  $-3$ ,  $0$ ,  $3$ ,  $6$  and  $9$  arcmin, corresponding to  $\Delta x/3$ ,  $0$ ,  $\Delta x/3$ ,  $2\Delta x/3$ ,  $\Delta x$ , respectively. The offsets for the third sub-frame are twice those values. The simulated stimulus was a  $1^\circ$ -wide white target traveling at  $6^\circ/\text{sec}$  on a black background. The capture rate was  $40\text{Hz}$ .

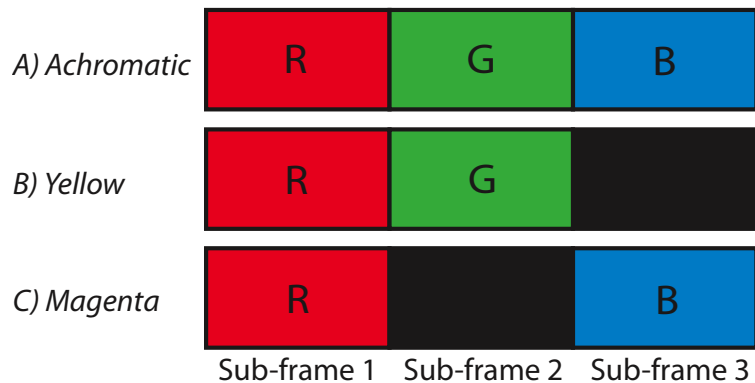


Figure 4.6: Protocols tested in the color-breakup experiment. The same sequential RGB protocol was used in all experiments, except that the blue or green channel was dropped in some conditions to create a yellow or magenta stimulus, respectively.

### 4.3 Psychophysical methods

#### Subjects

Three subjects, 22 to 32 years of age, participated. All had normal or corrected-to-normal visual acuity, stereo acuity, and color vision. Two were authors; one was not aware of the experimental hypotheses. Appropriate consent and debriefing were done according to the Declarations of Helsinki.

#### Apparatus

Stimuli were presented on a cathode-ray tube (CRT, ViewSonic G225f;  $1280 \times 960$  pixels, providing  $640 \times 960$  pixels per each-eye view) and viewed binocularly. Viewing distance was 107.3 cm where each pixel subtended 1arcmin. Refresh rate (or presentation rate) was 120Hz. The capture rate was 40Hz.

#### Stimulus and procedure

We tested the three presentation protocols depicted in Figure 4.6. All three decomposed color into red, green, and blue.

Yellow and magenta stimuli (Figure 4.6 B,C) were used to test whether an offset in the second sub-frame relative to the first had the same effect as twice that offset in the third sub-frame relative to the first. The yellow stimulus presented light during the first and second sub-frames (RGX) and the magenta stimulus presented light during the first and third (RXB).

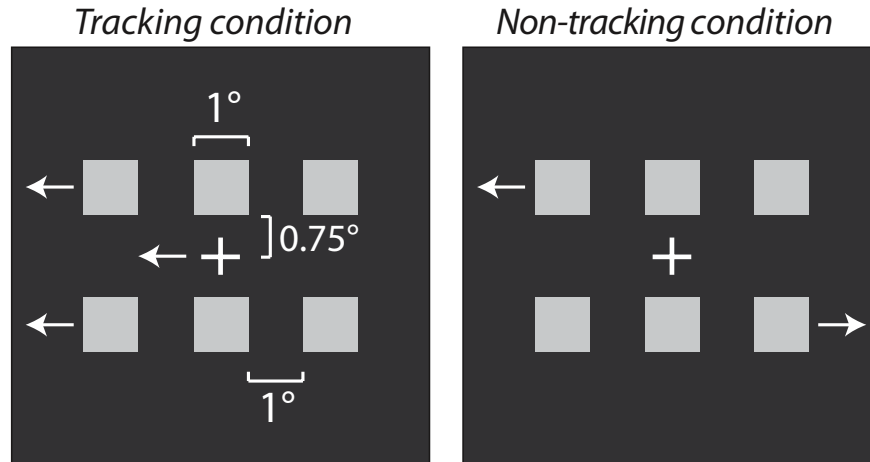


Figure 4.7: Stimulus used to measure color breakup. A series of bright rectangles moved across a dark background. In the tracking case, subjects tracked a cross that moved with the rectangles. In the non-tracking case, subjects fixated a stationary cross and the rectangles moved in opposite directions above and below the cross. In the figure, stimulus width is  $1^\circ$ , but the width varied between  $0.5$  and  $4^\circ$ .

We tested two eye-movement conditions. In the non-tracking condition, we presented five object speeds ( $2, 6, 10, 20,$  and  $30^\circ/\text{sec}$ , or  $1, 3, 5, 10,$  and  $15$  pixels/frame); in the tracking condition, we presented only three ( $2, 6, 10^\circ/\text{sec}$ ) because tracking becomes inaccurate at speeds greater than  $10^\circ/\text{sec}$  Robinson, Gordon, and Gordon, 1986.

The stimulus consisted of two groups of bright rectangles on a dark background (Figure 4.7). Each rectangle was  $1^\circ$  high. The widths were  $0.5, 1.0, 2.0,$  or  $4.0^\circ$ . The distance between adjacent rectangles was increased in proportion to stimulus speed so that no spatio-temporal aliasing occurred. In the tracking condition, the two groups of rectangles traveled horizontally at equal speed in the same direction. A fixation cross, located between the two groups, moved with the rectangles. To provide enough time for the subject to fixate, we presented the cross at its initial position for  $0.5\text{sec}$  before the onset of the rectangles. When the moving rectangles appeared, the cross moved with them. The rectangles stimulus was presented for  $1\text{sec}$ . In the non-tracking condition, the cross was presented midway between the rows of rectangles and was stationary. The two groups of rectangles travelled in opposite directions. By presenting opposing motions, we made it easier for subjects to hold the eyes stationary. In pilot testing, we noticed transient color breakup at the beginning and end of some presentations, probably due to reflexive tracking eye movements. To minimize such tracking, we faded the rectangles in and out at the beginning and end of each trial, for the non-tracking condition only. The fade in and fade out was not included in the model because the sensation of color breakup is strongest when the stimulus is brightest. After each presentation, subjects indicated with a key press whether or not they had perceived

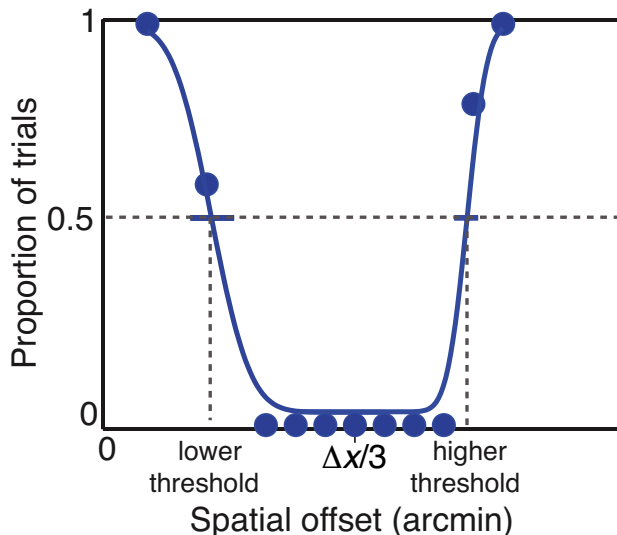


Figure 4.8: Psychometric function for perception of color breakup. Spatial offsets were added to the second and third sub-frames as described in the text. The proportion of trials in which color breakup was reported is plotted as a function of the offset of the second sub-frame. The data are from the achromatic condition with a stimulus speed of  $10^\circ/\text{sec}$  and stationary fixation. The value of the offset that should eliminate breakup is  $\Delta x/3$  (Equation 4.14). The curves are two cumulative Gaussians fit to the data using a maximum-likelihood criterion.

color breakup. They were instructed to respond positively only when they perceived hue variation at the edges. Thus, they did not respond if they perceived other motion artifacts such as judder or edge banding Hoffman et al., 2011.

We examined how spatial offsets added to the second and third sub-frames affect breakup visibility. One offset was added to the green component, and twice that offset to the blue. For each condition, we presented nine different offsets according to the method of constant stimuli. For a tracked stimulus moving at speed  $v$ , the offset  $\delta$  for the second sub-frame that should yield no breakup is:

$$\delta = \frac{v}{f}, \quad (4.14)$$

where  $f$  is the presentation rate. The predicted offset for the third sub-frame is  $2\delta$ . Using the same logic, we predict that the nulling offset for a yellow stimulus (in which only the first and second sub-frames are illuminated) would be  $\delta$  added to the second sub-frame, and that the nulling offset for magenta (in which the first and third sub-frames are illuminated) would be  $2\delta$  added to the third sub-frame. Expressed as proportions of the displacement  $\Delta x$  per frame,  $\delta$  and  $2\delta$  correspond to  $\Delta x/3$  and  $2\Delta x/3$ , respectively.

Figure 4.8 illustrates typical data for one speed and presentation protocol. The propor-

tion of trials in which color breakup was reported is plotted as a function of the spatial offset. We fit the data with two cumulative Gaussians, one on each side, using a maximum-likelihood criterion Fründ et al., 2011; Wichmann and Hill, 2001b. In all conditions and with all observers, there was a spatial offset that resulted in color breakup being reported on fewer than 20% of the trials, so we were able to reliably locate the descending and ascending portions of the psychometric data.

If we had tested all conditions, the experiment would have consisted of 10,800 trials per subject: 3 presentation protocols  $\times$  2 eye-movement conditions  $\times$  4 stimulus widths  $\times$  5 speeds  $\times$  9 offsets  $\times$  10 trials. Fewer trials were actually presented because some of those combinations of conditions were not realizable. About 5 hours were required for each subject to complete the whole experiment.

## 4.4 Results

For each stimulus speed, we found the range of spatial offsets that yielded no or few reports of color breakup. The data figures indicate those ranges with shaded regions. The black lines at the edges of the shaded regions are the offsets that yielded reports of breakup on half the trials.

Figure 4.9 shows the results with the yellow and magenta stimuli. Recall that the first and second sub-frames were illuminated for the yellow stimulus (RGX) and the first and third for the magenta stimulus (RXB). These stimuli were used to test our assumption that a spatial offset of  $\delta$  in the second sub-frame had the same effect on breakup as an offset of  $2\delta$  in the third sub-frame. Subjects tracked the stimulus during this experiment. The yellow and magenta shaded regions represent the offsets for which viewers perceived color breakup on fewer than half the trials for the yellow and magenta stimuli, respectively; the gray regions represent offsets for which viewers perceived breakup less than half the time for both stimuli. The dashed lines represent the predicted offsets required to eliminate breakup from Equation 4.14. As you can see, the experimental results are quite consistent with the predictions from that equation. In particular, the offset that eliminated breakup with the magenta stimulus was, as predicted, twice the offset that eliminated breakup with the yellow stimulus. In all subsequent figures, we will not plot the data for the third sub-frame because it is redundant.

Having confirmed that an offset of  $2\delta$  in the third sub-frame has an equivalent effect to an offset of  $\delta$  in the second sub-frame, we proceeded with the main experiment with bright achromatic stimuli (i.e., R, G, and B illuminated in each frame). For each stimulus speed, we found how often color breakup was reported for different offsets of the second and third sub-frames relative to the first. Figure 4.10 shows the results for the individual subjects. The shaded regions represent the offsets for which viewers perceived color breakup on fewer than half the trials. As predicted, breakup was minimized in the tracking condition when the offset was  $\Delta x/3$ . That offset also minimized breakup in the non-tracking condition when the speed was  $10^\circ/\text{sec}$  or slower. At faster speeds, the offset that minimized breakup was



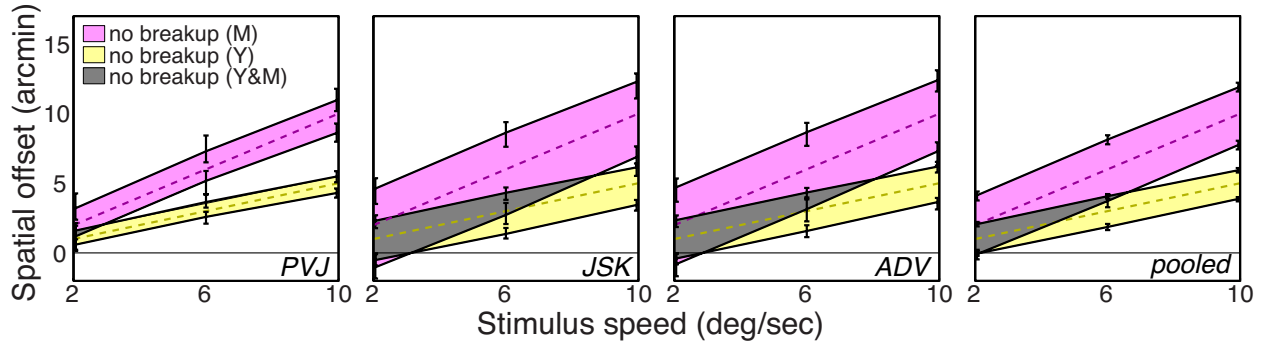


Figure 4.9: Color breakup when subjects tracked a yellow or magenta stimulus. Spatial offsets are plotted for different stimulus speeds. The shaded regions indicate the combinations of offsets and speeds that yielded breakup on fewer than half the trials. The first three panels show individual subject data, and the last panel shows data averaged across subjects. The yellow and magenta dashed lines correspond to the predicted nulling offsets for yellow and magenta, respectively (Equation 4.14). The yellow region corresponds to the range of offsets that produced color breakup on fewer than half the trials for the yellow stimulus, while the magenta region corresponds to the range of offsets that produced breakup on fewer than half the trials for the magenta stimulus. The gray region is the range of offsets that produced no color breakup for both stimuli. Error bars represent 95% confidence intervals.

smaller than  $\Delta x/3$ . The upper row of Figure 4.11 shows the data averaged across subjects. The lower panels show the predictions from the model. The ranges of offsets that minimize breakup according to the model were quite similar to the range we observed empirically. They were essentially identical in the tracking condition and were quite similar in the non-tracking condition. Thus, the model seems to be an accurate predictor of the visibility of color breakup.

We observed in model simulations that the size of the moving stimulus had little effect on breakup visibility whether the eyes were moving or not. We investigated whether these predictions are consistent with viewers' percepts by presenting different stimulus widths (0.5, 1.0, 2.0, 4.0°) and measuring breakup with stationary fixation. The upper row of Figure 4.12 shows those experimental results and the lower row the model predictions. As you can see, there was very little effect of stimulus size, confirming the model's predictions.

There is one condition for which the model's predictions and experimental behavior differ noticeably: where the stimulus width was 0.5° and speed was 20°/sec. In that condition, the model predicts that breakup will be much less salient than we observed experimentally. That condition corresponds to stimulus movement such that  $\Delta x$  (the distance moved from one frame to the next) equals the object's width. In that special case, the luminances of R, G, and B remain constant but shifted spatially by  $\Delta x$  in each frame. With temporal low-pass filtering, each color component would be almost uniformly blurred across time, resulting in

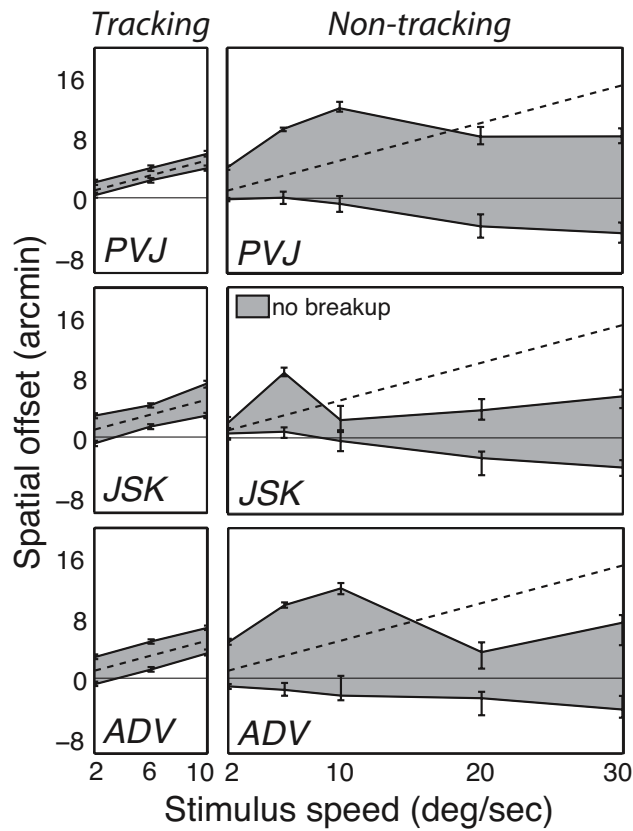


Figure 4.10: Color breakup as a function of stimulus speed and spatial offset of the second sub-frame relative to the first. The offset added to the third sub-frame was twice that of the second sub-frame. The left column shows the data for the tracking condition and the right column the data for the non-tracking condition. Each row represents the data from a different subject. The gray regions represent the offsets that yielded breakup on fewer than half the trials. The dashed lines represent the predictions of Eqn. 4.14. Error bars represent 95% confidence intervals.

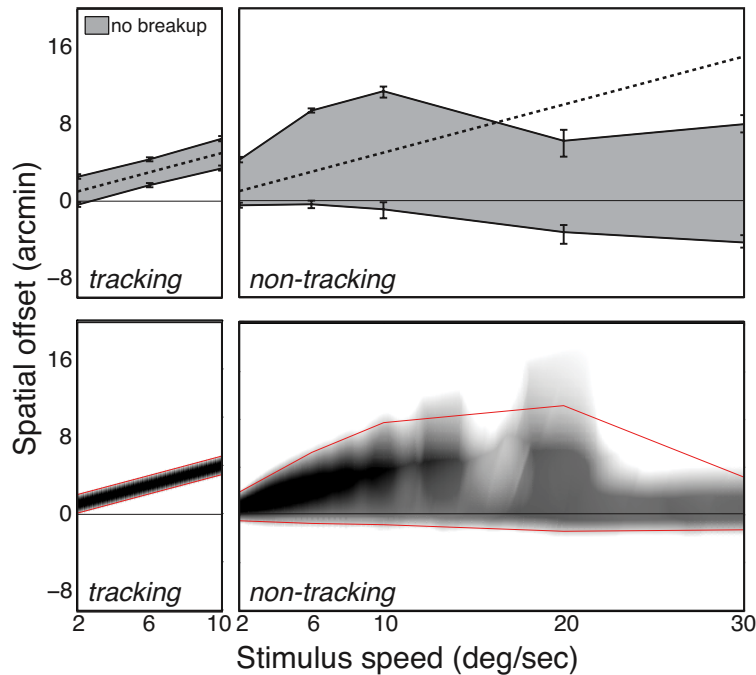


Figure 4.11: Top: Color-breakup data averaged across subjects. The spatial offset of the second sub-frame relative to the first is plotted as a function of stimulus speed. Left: subjects tracked the stimulus. Right: subjects maintained stationary fixation while the stimulus moved. The gray region represents the values of the spatial offset that yielded reports of color breakup on fewer than half the trials. Dashed lines represent the predictions of Eqn. 4.14. Error bars represent 95% confidence intervals. Bottom: Output of the color-breakup model. The dark regions correspond to the range of spatial offsets predicted to minimize breakup. The red contours represent the combinations of offsets and speeds that yielded a criterion amount of modulation at the speeds tested in the experiment.

less salient breakup. However, small fixational eye movements (jitter, drift and re-fixation saccades) are ever-present, and would cause sub-frames to be displaced slightly on the retina thereby avoiding the overlap that occurs with no eye movement. We examined the possibility that subjects' fixational movements caused breakup by incorporating such movements in the model. We found that these movements can indeed cause an increase in predicted breakup for that special condition. We did not, however, incorporate fixational movements into the model because their properties are difficult to measure and because their effect is probably only present in the special condition in which  $\Delta x$  equals the stimulus width.

The retinal eccentricity of the stimulus ranged from 1-10°. We wondered whether breakup was more visible in the peripheral or central visual field. One might argue that it would be most noticeable in the peripheral field where high temporal frequencies are more visible.

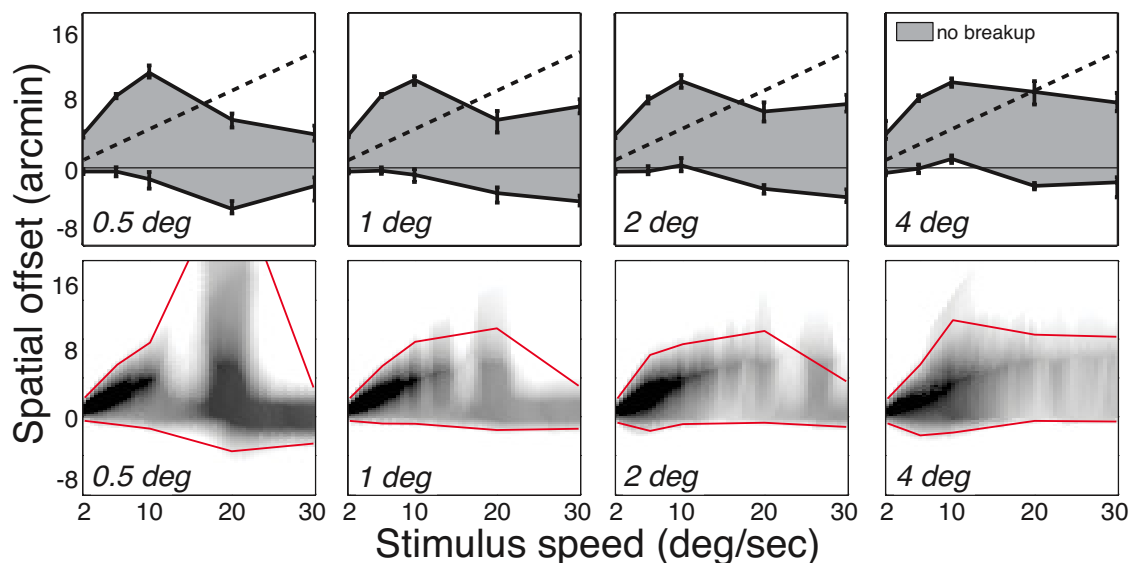


Figure 4.12: Top: Color breakup for four different stimulus widths (0.5, 1.0, 2.0, 4.0°), non-tracking condition, pooled across subjects. The gray region represents the range of spatial offsets that yielded reports of color breakup on fewer than half the trials. Bottom: Color-breakup predictions for the same four stimulus widths. The dark regions correspond to the range of spatial offsets predicted to minimize color breakup. The red contours represent the combinations of offsets and speeds that yielded a criterion amount of modulation at the speeds tested in the experiment.

But one might also argue that breakup would be most noticeable near the fovea where spatial resolution and sensitivity to hue variation are greatest. All subjects reported that breakup was most visible when the rectangles passed nearest the fovea suggesting that the poor resolution and reduced sensitivity of the retinal periphery make color breakup less noticeable in that part of the visual field.

## 4.5 Discussion

When a moving stimulus is presented by a color-sequential projector, color breakup is often observed. Consideration of the spatio-temporal filtering in color-opponent channels leads to the prediction that inserting specific spatial offsets to the second and third sub-frames will minimize or even eliminate breakup whether the viewer is tracking the stimulus or not. We confirmed this prediction in psychophysical experiments.

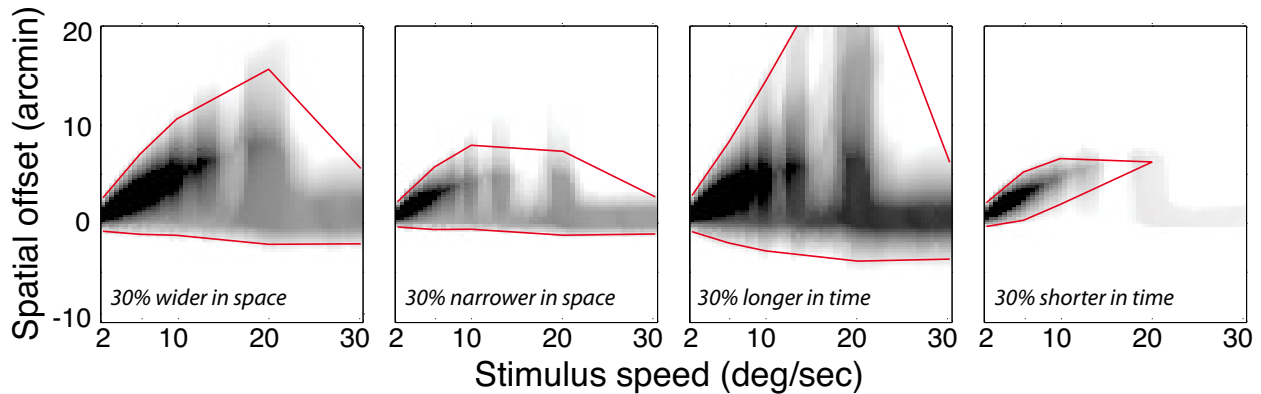


Figure 4.13: Properties of the impulse-response function and visibility of color breakup. The predicted visibility of color breakup from the output of the model is plotted for different spatial offsets and stimulus speeds, for the non-tracking condition. The stimulus was moving at  $6^\circ/\text{sec}$ . We varied the spatial width and temporal duration of the impulse-response function (Figure 4.4) by  $\pm 30\%$  to generate four different functions. From left to right, the panels show the results for impulse-response functions that are 30% wider in space, 30% narrower in space, 30% longer in time, and 30% shorter in time. The red contours represent the combinations of offsets and speeds that yielded a criterion amount of modulation at the speeds tested in the experiment.

### Shape of the impulse-response function

The similarity of the predictions and empirical observations supports the hypothesis that the visibility of color breakup is dependent on spatio-temporal filtering in early vision. We asked whether the similarity of predictions and results depends on specific filtering parameters or is more robust than that. To do this, we varied the spatial and temporal properties (width and duration) of the impulse-response function by  $\pm 30\%$ , yielding four different functions. We then convolved those functions with the same input stimuli and re-ran the simulations. The main results were unaffected by these variations: A spatial offset of  $\Delta x/3$  always minimized color breakup with tracking and non-tracking, and breakup saliency was always more dependent on spatial offset when the viewer was tracking than when fixation was stationary. Therefore, slightly different spatio-temporal impulse-response functions yield behavior that is consistent with our main conclusions. However, the quantitative agreement between the predictions and experimental observations is clearly dependent on the parameters of the assumed impulse-response function. This is illustrated in Figure 4.13, which shows some of the non-tracking results. The individual panels show the results for different impulse-response functions. The quantitative agreement between predictions and observations was best for the parameters derived from Burbeck and Kelly Burbeck and Kelly, 1980.

## Pursuit gain

When people track an object with a smooth eye movement, the velocity of the eye rotation does not always match the velocity of the object, so retinal slip occurs. Such errors of smooth movements can be quantified by the pursuit gain. A gain of 1 means of course that the eye speed is perfectly matched to the object speed. With predictable object motion, as in our experiments, the gain is typically close to 1 for object speeds up to  $10^\circ/\text{sec}$ ; it falls at greater speeds Robinson et al., 1986. In modeling our experimental data, we assumed a gain of 1 for the tracking condition because we did not present object speeds greater than  $10^\circ/\text{sec}$ .

## Saccades and breakup

Saccades are rapid, ballistic eye movements in which the eye rotates  $100\text{-}500^\circ/\text{sec}$  Carpenter, 1991. Such rapid movements should create quite salient color breakup even when the stimulus is stationary. Yohso and Ukai Yohso and Ukai, 2006 examined the effect of saccades of up to  $300^\circ/\text{sec}$  and found that the width of the color breakup percept was narrower than predicted based simply on the retinal loci of stimulation. The reduced visibility is almost certainly due to saccadic suppression, the reduction in sensitivity to incoming visual information during a saccade Matin, 1974. Although breakup is somewhat less visible during saccades relative to expectation, such eye movements remain an issue in color-sequential displays, particularly with high-contrast content. The problem is that saccades come at unpredictable times with unpredictable magnitudes and directions, so one cannot predict and compensate for them.

## Triple capture to create spatial offsets

When an object's position changes by  $\Delta x$  every frame, adding spatial offsets of  $\Delta x/3$  and  $2\Delta x/3$  to every second and third sub-frame minimizes or even eliminates color breakup. In realistic video content, however, the value of  $\Delta x$  may vary from one object to another. Thus, different offsets may be needed for different objects depending on their speeds.

One could in principle create the required offsets for each speed by using motion-compensated, frame-rate conversion (MCFRC) Kang, Cho, and Kim, 2007. MCFRC algorithms estimate motion between existing frames and create appropriate positions for intermediate frames. Thus, one could use such an algorithm to interpolate the appropriate second and third sub-frames and then use those interpolated frames to create the desired offsets. But MCFRC algorithms are computationally expensive and prone to artifacts when run in real-time Lee and Nguyen, 2010.

A simpler approach for creating the appropriate offsets would be to triple the capture rate. Figure 4.14 schematizes the idea. Instead of capturing content at  $f$  and presenting at  $3f$  (three sub-frames per frame), one would capture at  $3f$  and present the appropriate colors in each of the three sub-frames. By tripling the capture rate, one would generate the first, second, and third sub-frames from the captured data and could use those frames to

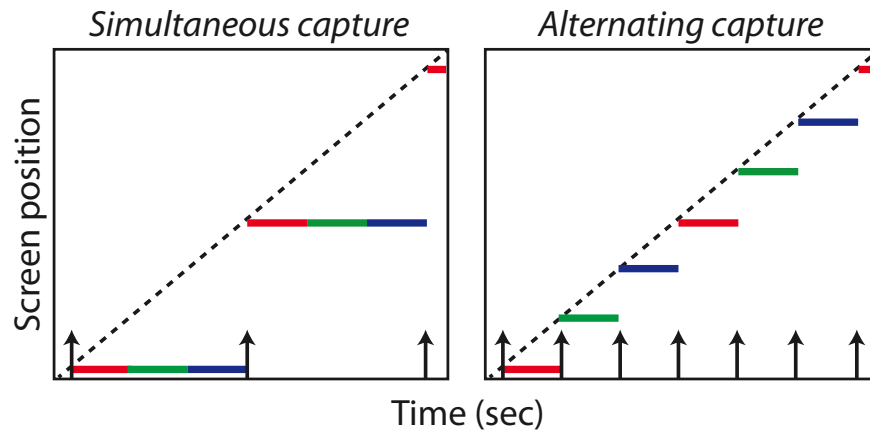


Figure 4.14: Simultaneous vs. alternating capture at triple the rate. Left: In simultaneous capture, the red, green, and blue components of the image are captured at the same time (black arrows), and presented sequentially. Right: In alternating capture, the red, green, and blue components are captured alternately over time and presented sequentially.

implement the appropriate spatial offsets for every speed. Thus, frames 1, 4, and 7 would be sent to the red channel, 2, 5, and 8 to the green channel, and 3, 6, and 9 to the blue channel. The desired spatial offsets would be baked into the content so no further manipulation would be needed.

This approach of capturing sub-frames at their appropriate spatial locations was used a long time ago in television. The first color television, demonstrated by Baird in 1928, used color-sequential presentation with a rotating color filter Apparatus for transmitting views or images to a distance, 1929. The filter rotated in synchrony with the content such that the capture and presentation rates were identical. This method in principle could have eliminated color breakup, but the early color-sequential systems had frame rates that were too low to do so Goldmark, Dyer, and Piore, 1942. The sequential technique for color presentation became obsolete when simultaneous color displays were introduced in 1947 Kell, 1947.

## Chapter 5

# Conclusion

In order to engineer better displays it is becoming increasingly important to consider the human element, particularly for stereoscopic displays, which have a unique set of constraints in order to provide different images for each eye. Optimizing visual experience for displays requires an understanding of the limitations of the visual system in time and space and knowing what tradeoffs can be made without sacrificing image quality. In this thesis I have discussed many of the perceptual artifacts that can occur in (mostly stereo) displays, and various methods to ameliorate them. Novel display techniques have the potential to improve viewing experience and reduce many of the artifacts associated with existing displays.

In Chapter 2, I demonstrated that a hybrid stereoscopic display protocol maintains the benefits of temporal and spatial interlacing methods while eliminating the drawbacks. In other words, this display has better spatial properties than spatial interlacing and better temporal properties than temporal interlacing. Motion appeared smoother, depth distortion was eliminated, flicker was reduced, and resolution was improved. Depending on the implementation, this display could also appear brighter than traditional temporally interlaced displays.

In Chapter 3, I investigated the effect of driving mode on perceptual artifacts on a 240Hz temporally interlaced stereoscopic 3D display. I demonstrated that the high frame rate of this display could be used to drive left- and right-eye views in unique ways that can reduce some of the artifacts associated with temporal interlacing, as well as create a viewing experience tailored to specific uses. If eliminating flicker is a priority, the display could be driven accordingly. Depth distortion can also be substantially reduced compared to a typical 120Hz temporally interlaced display. Additionally, the display can provide 4 views interlaced over time, allowing two people to experience their own full stereoscopic 3D content. High frame rates provide the freedom to customize a display based on the user's specific needs.

Chapter 4 discusses the phenomenon of color breakup in displays that present colors sequentially in time. I described a novel psychometric procedure used to measure the artifact, as well as a method for reducing the artifact under certain viewing conditions. I also described a method for modeling the saliency of color breakup using known properties of the human visual system. Results from the simulation agree with results from psychophysical



experiments, suggesting that adding a spatial offset to the content can reduce the perception of color breakup. I demonstrated that using alternating capture with triple the capture rate should allow for a substantial reduction in the visibility of color breakup, regardless of whether the viewer is tracking a stimulus or not.

# Bibliography

- 2020-1: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. *Internationale Telekommunikations-Union, Geneva, Switzerland.*
- 2022: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. *Internationale Telekommunikations-Union, Geneva, Switzerland.*
- 709-5: Parameter values for the HDTV standards for production and international programme exchange. *Internationale Telekommunikations-Union, Geneva, Switzerland.*
- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2), 284–299.
- Anzai, A., Ohzawa, I., & Freeman, R. D. (2001). Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect. *Nature Neuroscience*, 4(5), 513–518.
- Ardito, M. (1994). Studies of the influence of display size and picture brightness on the preferred viewing distance for HDTV programs. *SMPTE Motion Imaging Journal*, 103(8), 517–522.
- Arend, L., Lubin, J., Gille, J., Larimer, J., & Statler, I. C. (1994). Color breakup in sequentially-scanned LC displays. *NASA Technical Reports.*
- Backus, B. T., Banks, M. S., Van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, 39(6), 1143–1170.
- Bailey, I. L., & Lovie, J. E. (1976). New design principles for visual acuity letter charts. *American Journal of Optometry and Physiological Optics*, 53(11), 740–745.
- Baird, J. L. (1929). *Apparatus for transmitting views or images to a distance*. 1,699,270. 1929.
- Balram, N. (2008). *Methods and systems for improving low resolution and low frame rate video*. 12,033,490. 2008.
- Banks, M. S., Read, J. C. A., & Allison, R. S. (2012). Stereoscopy and the human visual system. *SMPTE Motion Imaging Journal*, 121(4), 24–43.
- Bex, P. J., Edgar, G. K., & Smith, A. T. (1995). Multiple images appear when motion energy detection fails. *Journal of Experimental Psychology: Human Perception and Performance*, 21(2), 231.

- Biswas, M., Balram, N., & Pathak, B. (2012). *Systems and methods for a motion compensated picture rate converter*. 8,340,185. 2012.
- Boman, D. A., & McMahan, R. P. (2007). Virtual reality: how much immersion is enough? *Computer*, 40(7), 36–43.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10.
- Burbeck, C. A., & Kelly, D. H. (1980). Spatiotemporal characteristics of visual mechanisms: Excitatory-inhibitory model. *Journal of the Optical Society of America*, 70(9), 1121–1126.
- Burr, D. C., & Morrone, M. C. (1993). Impulse-response functions for chromatic and achromatic stimuli. *Journal of the Optical Society of America A*, 10(8), 1706–1713.
- Burr, D. C., & Ross, J. (1979). How does binocular delay give information about depth? *Vision Research*, 19(5), 523–532.
- Burr, D. C., Ross, J., & Morrone, M. C. (1986). Smooth and sampled motion. *Vision Research*, 26(4), 643–652.
- Byrn, J., Schluender, S., & Divino, C. (2007). Three-dimensional imaging improves surgical performance for both novice and experienced operators using the da Vinci robot system. *The American Journal of Surgery*, 193(4), 519–522.
- Campbell, F. W., & Green, D. G. (1965a). Monocular versus binocular visual acuity. *Nature*, 208, 191–192.
- Campbell, F. W., & Green, D. G. (1965b). Optical and retinal factors affecting visual resolution. *The Journal of Physiology*, 181(3), 576.
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology*, 197(3), 551.
- Carpenter, R. H. S. (1991). *Eye movements*. Ann Arbor, MI: CRC Press.
- Cavonius, C. R. (1979). Binocular interactions in flicker. *The Quarterly Journal of Experimental Psychology*, 31(2), 273–280.
- Chan, H. P., Goodsitt, M. M., Helvie, M. A., Hadjiiski, L. M., Lydick, J. T., Roubidoux, M. A., ... Sahiner, B. (2005). ROC study of the effect of stereoscopic imaging on assessment of breast lesions. *Medical Physics*, 32(4), 1001–1009.
- Cheng, Y. K., & Shieh, H. P. D. (2009). Relative contrast sensitivity for color break-up evaluation in field-sequential-color LCDs. *Display Technology*, 5(10), 379–384.
- Cooper, E. A., Burge, J., & Banks, M. S. (2011). The vertical horopter is not adaptable, but it may be adaptive. *Journal of Vision*, 11(3), 20.
- Cooper, E. A., Jiang, H., Vildavski, V., & Farrell, J. E. (2013). Assessment of OLED displays for vision research. *Journal of Vision*, 13(12), 16.
- Cowan, M., Greer, J., & Lipton, J. (2013). *Projection of stereoscopic images using linearly polarized light*. 8,400,492. 2013.
- Davis, J., Hsieh, Y.-H., & Lee, H.-C. (2015). Humans perceive flicker artifacts at 500Hz. *Nature Scientific Reports*, 5, 7861.
- Dawson, S. (2012). Active versus passive. *Connected Home Australia*, 46–48.

- de Witt, S. (2012). Active/passive 3D myth revisited. **retrieved from** <http://stijndewitt.wordpress.com/2012/03/03/active-vs-passive-3d-myth-revisited/>
- Didyk, P., Eisemann, E., Ritschel, T., Myszkowski, K., & Seidel, H.-P. (2010). Apparent display resolution enhancement for moving images. *ACM Transactions on Graphics (TOG)*, 29(4), 113.
- Elze, T. (2010). Achieving precise display timing in visual neuroscience experiments. *Journal of Neuroscience Methods*, 191(2), 171–179.
- Elze, T., & Tanner, T. G. (2012). Temporal properties of liquid crystal displays: Implications for vision science experiments. *PloS One*, 7(9), e44048.
- Elze, T., Taylor, C., & Bex, P. J. (2013). An evaluation of organic light emitting diode monitors for medical applications: Great timing, but luminance artifacts. *Medical Physics*, 40(9), 092701.
- Emoto, M., & Sugawara, M. (2014). Still images of high spatial resolution enable short TV viewing distances. *ITE Transactions on Media Technology and Applications*, 2(2), 185–191.
- Feng, X., Pan, H., & Daly, S. (2008). Comparisons of motion-blur assessment strategies for newly emergent LCD and backlight driving technologies. *Journal of the Society for Information Display*, 16(10), 981–988.
- Frohlich, B., Barrass, S., Zehner, B., Plate, J., & Gobel, M. (1999). Exploring geo-scientific data in virtual environments. *Proceedings of the Conference on Visualization '99: celebrating 10 years, IEEE Computer Society*, 169–173.
- Fründ, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, 11(6), 16.
- Fujine, T., Kikuchi, Y., & Sugino, M. (2007). Real-life in-home viewing conditions for flat panel displays and statistical characteristics of broadcast video signal. *Japanese Journal of Applied Physics*, 46(3S), 1358.
- Goldmark, P. C., Dyer, J. N., & Piore, E. R. (1942). Color Television-Part 1. In *Proceedings of the ire* (Vol. 30, 4, pp. 162–182).
- Hakala, J., Oittinen, P., & Häkkinen, J. (2015, in press). Depth artifacts caused by spatial interlacing in stereoscopic 3D displays. *ACM Trans. Appl. Percept.*
- Heesch, F. H., & Klompenhouwer, M. A. (2008). Characterizing displays by their temporal aperture: A theoretical framework. *Journal of the Society for Information Display*, 16(10), 1009–1019.
- Held, R. T., & Hui, T. T. (2011). A guide to stereoscopic 3D displays in medicine. *Academic Radiology*, 18(8), 1035–1048.
- Hernandez, A., Basset, O., Bremond, A., & Magnin, I. E. (1998). Stereoscopic visualization of three-dimensional ultrasonic data applied to breast tumours. *European Journal of Ultrasound*, 8(1), 51–65.
- Hoffman, D. M., Johnson, P. V., Kim, J., Vargas, A., & Banks, M. (2015, in press). 240Hz OLED technology properties that can enable improved image quality. *Journal of the Society for Information Display*.

- Hoffman, D. M., Karasev, V. I., & Banks, M. S. (2011). Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth. *Journal of the Society for Information Display*, 19(3), 271–297.
- Holliman, N. S., Dodgson, N. A., Favalora, G. E., & Pockett, L. (2011). Three-dimensional displays: A review and applications analysis. *Broadcasting, IEEE Transactions on*, 57(2), 362–371.
- Hong, J., Kim, Y., Choi, H.-J., Hahn, J., Park, J.-H., Kim, H., . . . Lee, B. (2011). Three-dimensional display technologies of recent interest: Principles, status, and issues. *Applied Optics*, 50(34), H87–H115.
- Hornbeck, L. J. (1997). Digital light processing (TM) for high-brightness, high-resolution applications. *Proc SPIE*, 3013, 27–40.
- Hu, Y. (2006). The role of three-dimensional visualization in surgical planning of treating lung cancer. *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 646–649.
- Johnson, P. V., Kim, J., & Banks, M. S. (2014). The visibility of color breakup and a means to reduce it. *Journal of Vision*, 14(4), 10.
- Johnson, P. V., Kim, J., Hoffman, D. M., & Banks, M. S. (2014). 55.1: Distinguished paper: Motion artifacts on 240Hz OLED stereoscopic 3D displays. In *SID symposium digest of technical papers* (Vol. 45, 1, pp. 797–800). The Society for Information Display. Blackwell Publishing Ltd.
- Johnson, P. V., Kim, J., Hoffman, D. M., Vargas, A., & Banks, M. (2015, in press). Motion artifacts on 240Hz OLED stereoscopic 3D displays. *Journal of the Society for Information Display*.
- Jorke, H., & Fritz, M. (2003). Infitec-a new stereoscopic visualisation tool by wavelength multiplex imaging. *Proceedings of Electronic Displays*.
- Jorke, H., Simon, A., & Fritz, M. (2009). Advanced stereo projection using interference filters. *Journal of the Society for Information Display*, 17(5), 407–410.
- Kane, D., Guan, P., & Banks, M. S. (2014). The limits of human stereopsis in space and time. *The Journal of Neuroscience*, 34(4), 1397–1408.
- Kang, S. J., Cho, K. R., & Kim, Y. H. (2007). Motion compensated frame rate up-conversion using extended bilateral motion estimation. *Consumer Electronics, IEEE Transactions on*, 53(4), 1759–1767.
- Kell, R. D. (1947). An experimental simultaneous color-television system: Part I-Introduction. In *Proceedings of the ire* (Vol. 35, 9, pp. 861–862).
- Kelley, E. F. (2011). Frontline technology: resolving resolution. *Information Display*, 27(9), 18.
- Kelly, D. H. (1969). Flickering patterns and lateral inhibition. *Journal of the Optical Society of America*, 59(10), 1361–1368.
- Kelly, D. H. (1977). Visual contrast sensitivity. *Journal of Modern Optics*, 24(2), 107–129.
- Kelly, D. H. (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10), 1340–1349.

- Kelly, D. H. (1983). Spatiotemporal variation of chromatic and achromatic contrast thresholds. *Journal of the Optical Society of America*, 73(6), 742–749.
- Kim, J. S., & Banks, M. S. (2012). 65.1: Effective spatial resolution of temporally and spatially interlaced stereo 3D televisions. In *SID symposium digest of technical papers* (Vol. 43, 1, pp. 879–882). The Society for Information Display. Blackwell Publishing Ltd.
- Klompenerhouwer, M. A. (2006). 54.1: Comparison of LCD motion blur reduction methods using temporal impulse response and MPRT. In *SID symposium digest of technical papers* (Vol. 37, 1, pp. 1700–1703). The Society for Information Display. Blackwell Publishing Ltd.
- Koenderink, J. J., Bouman, M. A., de Mesquita, A. E. B., & Slappendel, S. (1978). Perimetry of contrast detection thresholds of moving spatial sine wave patterns. II. The far peripheral visual field (eccentricity 0–50). *Journal of the Optical Society of America*, 68(6).
- Kuroki, Y. (2012). Improvement of 3D visual image quality by using high frame rate. *Journal of the Society for Information Display*, 20(10), 566–574.
- Lambooi, M., Fortuin, M., Heynderickx, I., & IJsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53(3), 30201–30214.
- Lee, Y.-L., & Nguyen, T. (2010). High frame rate motion compensated frame interpolation in high-definition video processing. *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*.
- Lipton, L. (1982). *Foundations of the stereoscopic cinema: a study in depth*. New York, NY: Van Nostrand Reinhold.
- Marks, P. (2012). The switch to high frame-rate films may not be a smooth one. *New Scientist*, 214(2863), 20.
- Mason, S. J., & Zimmerman, H. J. (1960). *Electronic circuits, signals and systems*. New York, NY: Wiley.
- Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, 81(12), 899.
- Mendiburu, B. (2009). *3D movie making: Stereoscopic digital cinema from script to screen*. Oxford, UK: Focal Press, Elsevier.
- Miseli, J. (2006). Taking on motion artifacts evaluation in the VESA FPDM. *Journal of the Society for Information Display*, 14(11), 987–997.
- Moll, T., Douek, P., Finet, G., Turjman, F., Picard, C., Revel, D., ... (1998). Clinical assessment of a new stereoscopic digital angiography system. *Cardiovascular and Interventional Radiology*, 21(1), 11–16.
- Morgan, M. J., & Thompson, P. (1975). Apparent motion and the Pulfrich effect. *Perception*, 4(1), 3–18.
- Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1), 381–400.

- Nelson, T. R., Ji, E. K., Lee, J. H., Bailey, M. J., & Pretorius, D. H. (2008). Stereoscopic evaluation of fetal bony structures. *Journal of Ultrasound in Medicine*, 27(1), 15–24.
- Park, M., Kim, J., & Choi, H.-J. (2014). Effect of interlacing methods of stereoscopic displays on perceived image quality. *Applied Optics*, 53(3), 520–527.
- Pastoor, S. (2005). 3D displays. In O. Schreer, P. Kauff & T. Sikora (Eds.), *3D video communication*. Chichester, England: John Wiley and Sons.
- Patel, H. R., Ribal, M. J., Arya, M., Nauth-Misir, R., & Joseph, J. V. (2007). Is it worth revisiting laparoscopic three-dimensional visualization? A validated assessment. *Urology*, 70(1), 47–49.
- Patterson, R. E. (2013). Interactive stereoscopic displays. *Journal of the Society for Information Display*, 21(3), 142–150.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Poirson, A. B., & Wandell, B. A. (1993). Appearance of colored patterns: Pattern-color separability. *Journal of the Optical Society of America A*, 10(12), 2458–2470.
- Post, D. L., Monnier, P., & Calhoun, C. S. (1997). Predicting color breakup on field-sequential displays. *AeroSense'97*, 57–65.
- Post, D. L., Nagy, A. L., & Monnier, P. (1998). 38.3: Predicting Color Breakup on Field-Sequential Displays: Part 2. In *SID symposium digest of technical papers* (Vol. 29, 1, pp. 1037–1040). The Society for Information Display. Blackwell Publishing Ltd.
- Read, J. C. A., & Cumming, B. G. (2005). The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth. *Journal of Vision*, 5(5), 3.
- Robinson, D. A., Gordon, J. L., & Gordon, S. E. (1986). A model of the smooth pursuit eye movement system. *Biological Cybernetics*, 55(1), 43–57.
- Rosenthal, M., State, A., Lee, J., Hirota, G., Ackerman, J., Keller, K., . . . Fuchs, H. (2002). Augmented reality guidance for needle biopsies: an initial randomized, controlled trial in phantoms. *Medical Image Analysis*, 6(3), 313–320.
- Schreiber, K. M., Hillis, J. M., Filippini, H. R., Schor, C. M., & Banks, M. S. (2008). The surface of the empirical horopter. *Journal of Vision*, 8(3), 7.
- Sechrist, S. (2011). Display week 2011 review: 3-D. (Vol. 27, 7–8, pp. 16–18).
- Sharp, G. D., & Robinson, M. G. (2007). Enabling stereoscopic 3D technology. In *Electronic imaging 2007* (pp. 64900X–64900X). International Society for Optics and Photonics.
- Someya, J., & Sugiura, H. (2007). Evaluation of liquidcrystal display motion blur with moving picture response time and human perception. *Journal of the Society for Information Display*, 15(1), 79–86.
- Soneira, R. M. (2011). 3D TV display technology shoot-out. **retrieved from** [http://www.displaymate.com/3D\\_TV\\_ShootOut\\_1.htm](http://www.displaymate.com/3D_TV_ShootOut_1.htm)
- Song, W., Teunissen, K., Li, X., & Zhang, Y. (2009). Evaluation of motion performance on scanning backlight LCDs. *Journal of the Society for Information Display*, 17(3), 251–261.
- Stork, D. G., & Falk, D. S. (1987). Temporal impulse responses from flicker sensitivities. *Journal of the Optical Society of America A*, 4(6), 1130–1135.

- Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, part I*, 757–764.
- Taffinder, N., Smith, S. G., Huber, J., Russell, R. C. G., & Darzi, A. (1999). The effect of a second-generation 3D endoscope on the laparoscopic precision of novices and experienced surgeons. *Surgical Endoscopy*, *13*(11), 1087–1092.
- Tong, X., Heeger, D. J., & van den Branden, L. C. (1999). Video quality evaluation using ST-CIELAB. *Electronic Imaging '99*, 185–196.
- Tyler, C. W. (1985). Analysis of visual modulation sensitivity. II. Peripheral retina and the role of photoreceptor dimensions. *Journal of the Optical Society of America A*, *2*(3), 393–398.
- Tyler, C. W. (1987). Analysis of visual modulation sensitivity. III. Meridional variations in peripheral flicker sensitivity. *JOSA A*, *4*(8), 1612–1619.
- Van Berkel, C., & Clarke, J. A. (1997). Characterization and optimization of 3D-LCD module design. In *Electronic imaging'97* (pp. 179–186). International Society for Optics and Photonics.
- Watson, A. B. (2010). Display motion blur: Comparison of measurement methods. *Journal of the Society for Information Display*, *18*(2), 179–190.
- Watson, A. B. (2013). High frame rates and human vision: A view through the window of visibility. *SMPTE Motion Imaging Journal*, *122*(2), 18–32.
- Watson, A. B., Ahumada, A. J., Jr, & Farrell, J. E. (1986). Window of visibility: A psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A*, *3*(3), 300–307.
- Watson, O. J. (2013). 30.2: Driving scheme required for blurfree motion of a target moving at 480 pps. In *SID symposium digest of technical papers* (Vol. 44, 1). The Society for Information Display. Blackwell Publishing Ltd.
- 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*, 371–394.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314–1329.
- Winkler, S. (2005). *Digital video quality: Vision models and metrics*. Chichester, England: John Wiley and Sons.
- Yohso, A., & Ukai, K. (2006). How color breakup occurs in the humanvisual system: The mechanism of the color breakup phenomenon. *Journal of the Society for Information Display*, *14*(12), 1127–1133.
- Yun, J. D., Kwak, Y., & Yang, S. (2013). Evaluation of perceptual resolution and crosstalk in stereoscopic displays. *Journal of Display Technology*, *9*(2), 106–111.
- Zhang, X., & Farrell, J. E. (2003). Sequential color breakup measured with induced saccades. *Proc SPIE*, *5007*, 210–217.



- Zhang, X., & Wandell, B. A. (1997). A spatial extension of CIELAB for digital colorimage reproduction. *Journal of the Society for Information Display*, 5(1), 61–63.