



UNIVERSITY OF  
CAMBRIDGE

# Perceptual models for high-refresh-rate rendering

György Dénes



Gonville and Caius College

*This dissertation is submitted on 23rd January 2020 for  
the degree of Doctor of Philosophy*





---

# DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or am concurrently submitting, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation does not exceed the prescribed limit of 60 000 words.

György Dénes  
January, 2020

*“Time is the most valuable thing that a man can spend”*

*Theophrastus*

---

# ABSTRACT

**György Dénes: Perceptual models for high-refresh-rate rendering**

Rendering realistic images requires substantial computational power. With new high-refresh-rate displays as well as the renaissance of virtual reality (VR) and augmented reality (AR), one cannot expect that GPU performance will scale fast enough to meet the requirements of immersive photo-realistic rendering with current rendering techniques.

In this dissertation, I follow the dual of the well-known computer vision approach: *vision is inverse graphics*: to improve graphical algorithms, I consider the operation of the human visual system. I propose to model and exploit the limitations of the visual system in the context of novel high-refresh-rate displays; specifically, I focus on spatio-temporal perception, a topic that has received remarkably less attention than spatial-only perception so far.

I present three main contributions. First, I demonstrate the validity of the perceptual approach by presenting a conceptually simple rendering technique motivated by our eyes' limited sensitivity to high spatio-temporal change which reduces the rendering load and transmission requirement of current-generation VR headsets without introducing perceivable visual artefacts. Second, I present two visual models related to motion perception: (a) a metric for detecting flicker; and (b) a comprehensive visual model to predict perceived motion quality on monitors with arbitrary refresh rates and monitor resolutions. Third, I propose an adaptive rendering algorithm that utilises the proposed models. All algorithms operate on physical colorimetric units (instead of display-referenced pixel values), for which I provide the appropriate display measurements and models. All proposed algorithms and visual models are calibrated and validated with psychophysical experiments.



---

# ACKNOWLEDGEMENTS

With the realisation that most people rarely read the Acknowledgements, I will attempt to keep this passage reasonably short. The difficulty lies in the high number of people who deserve to be mentioned, which, if nothing else, might edify a naïve reader just how much unrealised support goes into the completion of a PhD programme.

First and foremost, I am thankful to my supervisor, Dr Rafał K. Mantiuk, who has provided guidance and encouragement throughout the past three and a half years. His interest in human vision inspired me to shift my focus to the exciting inter-disciplinary field of perceptual graphics. I also owe a debt of gratitude to Professor Peter Robinson, who advised me and helped me navigate the first year of my PhD. Even though my research interest has diverged from his, I could not have wished for a better mentor in the first few months of my research. I also wish to express my best thanks to my undergraduate director of studies, Dr Alastair Beresford, who continues to demonstrate inexhaustible patience for writing references and giving friendly advice over a cup of tea. I am also grateful to my first-year assessor, Professor Alan Blackwell, and my final examiners, Professor John Daugman and Dr Andrew Watson, for their insightful feedback and for organising the viva voce exam remotely during the time of the COVID-19 lockdown.

I would like to thank all my good friends and colleagues in Cambridge. Collaborations in the Computer Lab were often instructive and certainly fun. In particular, I would like to extend my gratitude to Aliaksei Mikhailiuk, Kuba Maruszczyk, Akshay Jindal, Fangcheng Zhong, and Nanyang Ye. While the Department of Computer Science and Technology is full of amazing individuals, I consider myself lucky to have been influenced by people from different fields as well. I am grateful for the continuous support of both Andrea Kocsis and Tiffany Ki, who have always expressed a curious amount of interest towards my research, helped me think outside the box, and seemed keen to contemplate with me the big questions in Life.

I also wish to thank Marcell Szmandray for permitting me to use his excellent panorama

pictures for my psychophysical experiments. I greatly appreciated his organising so many mountain hikes, taking all these great photographs, then permitting me to recycle his work in subsequent publications and in this dissertation.

Although one rarely considers putting a price tag on a PhD, I consider myself lucky to have received both a stipend and funding for my tuition fee from the Engineering and Physical Research Council (EPSRC). Without their help I would not have been able to start this programme.

Life as an academic researcher has proved to be quite a challenging adventure; one which I could not have persevered in if not for the opportunity to take a break from time to time and pursue my dream to teach. I am indebted to all directors of studies who trusted me to supervise their students in the Department of Computer Science and Technology for Part II projects and for courses ranging from Graphics to Prolog and Security. In my third year, I also enjoyed the opportunity to volunteer in Ms Rosa Sanches's Year 4 class in Fawcett Primary as part of the STIMULUS programme. I will always think fondly of the Monday mornings I spent with them. Finally, I am most grateful for the chance to repeatedly volunteer at my "alma mater" secondary school. I cannot thank Mrs Ildiko Lutter, Mr Andras Lutter, Ms Katalin Veres, Mrs Kati Zentai and all other teachers of Dunakeszi Radnoti Miklos Gimnazium enough for welcoming me into their community.

Last but not least, I would never have been able to get this far without the support, kindness and love I've received both from my family and all my friends. I hope that those named above can derive at least some joy, and those accidentally missed can forgive my carelessness.

---

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>15</b>
1.1	Motivation . . . . .	15
1.2	Hypothesis . . . . .	16
1.3	Structure of the dissertation . . . . .	16
1.4	Publications, presentations, awards . . . . .	17
<b>2</b>	<b>Background on spatio-temporal perception</b>	<b>21</b>
2.1	Overview of human visual system . . . . .	22
2.2	Early vision . . . . .	22
2.3	Visual pathways . . . . .	23
2.4	Contrast sensitivity . . . . .	24
2.4.1	Contrast definition . . . . .	24
2.4.2	Spatial contrast sensitivity function . . . . .	25
2.4.3	Temporal contrast sensitivity . . . . .	26
2.5	Temporal integration . . . . .	27
2.6	Motion perception . . . . .	28
2.6.1	Eye motion . . . . .	28
2.6.2	Apparent motion . . . . .	30
2.6.3	Motion artefacts . . . . .	31
2.6.3.1	Flicker . . . . .	31
2.6.3.2	False edges . . . . .	31
2.6.3.3	Blur . . . . .	34
2.6.3.4	Judder . . . . .	34
2.6.4	Motion sharpening . . . . .	35
2.7	Summary . . . . .	35

<b>3</b>	<b>Related work</b>	<b>37</b>
3.1	Critical resolution . . . . .	37
3.1.1	Spatial resolution . . . . .	37
3.1.2	Temporal resolution . . . . .	38
3.2	Temporal display technologies . . . . .	40
3.2.1	Blur reduction . . . . .	40
3.2.2	Display sync (V-Sync, G-Sync, Free-Sync) . . . . .	41
3.3	Multiplexing techniques . . . . .	42
3.3.1	Resolution vs. colour . . . . .	42
3.3.2	Resolution vs. refresh rate . . . . .	43
3.3.3	Temporal multiplexing . . . . .	44
3.3.3.1	Resolution vs. time . . . . .	45
3.3.3.2	Colour vs. time . . . . .	45
3.3.3.3	Temporal coherence . . . . .	45
3.3.4	Non-uniform (foveated) rendering . . . . .	47
3.4	Image and video metrics . . . . .	47
3.5	Summary . . . . .	48
<b>4</b>	<b>Display profiling</b>	<b>49</b>
4.1	Displays . . . . .	49
4.2	Measurement equipments, methods . . . . .	50
4.2.1	Photometric measurements . . . . .	50
4.2.2	Temporal measurements . . . . .	51
4.3	LCD displays . . . . .	51
4.3.1	Backlight . . . . .	53
4.3.2	Display model . . . . .	54
4.3.3	Overdrive . . . . .	55
4.4	OLED displays . . . . .	55
4.5	High-refresh-rate LCD model . . . . .	56
4.6	Summary . . . . .	59
<b>5</b>	<b>Exploiting perceptual insights: Temporal resolution multiplexing</b>	<b>61</b>
5.1	Introduction . . . . .	62
5.2	Method . . . . .	62
5.2.1	Frame integration . . . . .	64
5.2.2	Overshoots and undershoots . . . . .	66
5.2.3	Phase distortions . . . . .	67
5.2.4	Resolution reduction vs refresh rate (Experiment 5.1) . . . . .	68
5.3	Comparison with other techniques . . . . .	71



5.4	Applications . . . . .	74
5.4.1	Virtual reality . . . . .	74
5.4.2	High-refresh-rate monitors . . . . .	77
5.4.3	Portable devices . . . . .	77
5.5	Validation . . . . .	77
5.5.1	Virtual Reality (Experiments 5.2 and 5.3) . . . . .	77
5.5.2	Validation for desktop setup (Experiment 5.4) . . . . .	80
5.6	Limitations . . . . .	83
5.7	Summary . . . . .	84
<b>6</b>	<b>Multi-scale visual models</b>	<b>85</b>
6.1	Multi-band models of vision . . . . .	86
6.2	Banding artefacts . . . . .	86
6.3	Model design . . . . .	88
6.4	Extension for chromatic banding . . . . .	91
6.5	Model predictions . . . . .	92
6.6	Summary . . . . .	94
<b>7</b>	<b>A visual model for flicker</b>	<b>95</b>
7.1	Model design . . . . .	95
7.2	Psychophysical calibration (Experiment 7.1) . . . . .	98
7.3	Application . . . . .	102
7.4	Summary . . . . .	103
<b>8</b>	<b>Visual model for blur and judder</b>	<b>105</b>
8.1	Measuring motion quality (Experiment 8.1) . . . . .	106
8.2	A perceptual model for motion quality . . . . .	110
8.2.1	Blur due to spatio-temporal resolution and eye motion . . . . .	111
8.2.2	Motion-parallel and orthogonal blur . . . . .	114
8.2.3	From $\sigma$ to quality . . . . .	115
8.2.4	Judder ( $Q_J$ ) . . . . .	116
8.3	Model calibration . . . . .	117
8.3.1	Retinal blur due to motion (Experiment 8.2) . . . . .	117
8.3.2	Judder parameters (Experiment 8.3) . . . . .	122
8.3.3	Fitting the quality predictions . . . . .	124
8.3.4	Comparison with the model of Chapiro et al. . . . .	124
8.3.5	Ablation study . . . . .	127
8.3.6	The effect of resolution . . . . .	127
8.4	Model application . . . . .	128

8.4.1	Real-time implementation . . . . .	128
8.4.2	Experiment 8.4: psychophysical validation . . . . .	131
8.5	Limitations . . . . .	133
8.6	Summary . . . . .	135
<b>9</b>	<b>Conclusion</b>	<b>137</b>
9.1	Contributions . . . . .	137
9.2	Future work . . . . .	138
9.3	Final remarks . . . . .	139
	<b>Bibliography</b>	<b>141</b>
<b>A</b>	<b>Display Profiles</b>	<b>157</b>
A.1	Dell Inspiron 17R 7720 3D . . . . .	158
A.2	Samsung SyncMaster . . . . .	159
A.3	ASUS PG279Q . . . . .	160
A.3.1	High-refresh-rate model parameters . . . . .	161
A.4	HTC Vive . . . . .	162
A.5	Oculus Rift . . . . .	163
A.6	Huawei Mate Pro 9 – normal mode . . . . .	164
A.7	Huawei Mate Pro 9 – VR mode . . . . .	165
<b>B</b>	<b>Flicker marking stimuli</b>	<b>167</b>
<b>C</b>	<b>Motion quality model predictions</b>	<b>171</b>

---

# GLOSSARY

**AR** augmented reality.

**ASW** Oculus's asynchronous spacewarp.

**CFF** critical flicker frequency.

**cpd** cycles per visual degree.

**CSF** (spatial) contrast sensitivity function.

**fovea** central region (2 visual degrees) of the retina.

**GPU** graphics processing unit.

**HVS** human visual system.

**ppd** pixels per visual degrees.

**rgb** red, green, and blue.

**SPEM** smooth pursuit eye motion.

**stCSF** spatio-temporal contrast sensitivity function.

**tCSF** temporal contrast sensitivity function.

**VDP** visual difference predictor.

**VR** virtual reality.



---

---

# CHAPTER 1

---

## INTRODUCTION

### 1.1 Motivation

Perfect realism to the extent of indistinguishability from the real world has been the Holy Grail of computer graphics for a long time. Many recent advances in both real-time rendering and display technology attempt to bring us closer to creating such realistic imagery. For instance, ever-increasing resolutions reveal fine details, binocular and multi-focal setups add a sense of depth, and better animation quality improves perceived motion. Recently, virtual reality (VR) and augmented reality (AR) have been in the focus of such technological improvements, achieving reduced viewer discomfort, now striving to achieve complete realism in a fully-simulated environment.

However, achieving the goal of perfect realism is still a long way away from current rendering and display technology. *How long away exactly though?* Traditionally, rendering is reduced to computing a sequence of static two-dimensional frames. Industry standards would suggest at least 60 pixels per visual degree with a field of view of 110 degrees, two eyes, and a refresh rate of 140 Hz. This yields over 12 billion pixels rendered and displayed every second as a reasonable minimum – compared to the 400 million pixels per second of the latest VR headsets.

The required bandwidth is further increased as we want to increase the dynamic range or bring accommodation depth cues on multiple focal planes. Real-time generation and display of this vast amount of data is seemingly impossible even for the next few generations of graphics hardware.

Fortunately, perceived quality does not simply depend on the number of pixels and the frequency of frames. When visual content is judged by subjective human observers, quality becomes a complex function of multiple inter-related factors, such as spatio-temporal

content, scene motion, eye motion, and luminance level among others. In this dissertation, I propose rendering algorithms derived from the limitations and models of the visual system. These algorithms depart from the traditional sequence of constant-resolution rendering.

The vast and complex fields of perceptual graphics and visual modelling have been in the focus of numerous previous researches, producing valuable contributions, such as visual difference metrics, and foveated rendering algorithms. Many of the spatial factors are well-understood, however, the temporal domain has received remarkably little attention so far. Therefore, in this work, I restrict my focus to unique solutions that high-refresh-rate monitors offer — high refresh rate in this context is defined as more than the 60 Hz update of standard desktop monitors. Such techniques are expected to dominate the real-time game industry in the upcoming years both for desktop and virtual-reality setups. Consequently, the proposed algorithms and models are primarily concerned with motion quality and spatio-temporal perception.

## 1.2 Hypothesis

In this dissertation, I hypothesise that understanding the spatio-temporal limitations of the visual system can be utilised to reduce the computational complexity of high-refresh-rate rendering algorithms. The eventual consumers of the generated visual content are humans, therefore I claim that insights into the limitations of the visual system can help to design algorithms that are computationally cheaper but perceptually indistinguishable from their naïve counterparts. Furthermore, I postulate that invertible models of the visual system can predict the trade-off between rendering dimensions such as resolution, refresh rate, bit-depth, etc., maximising perceived quality even under constrained rendering budgets.

In a sense, I suggest to use the dual of the computer-vision philosophy that *vision is inverse graphics* (VIG). Computer vision has successfully shown that utilising rendering engines can help to more accurately estimate camera position, reconstruct 3D objects, or to train neural networks. In this dissertation, I attempt to show that utilising visual models can help to render in a more performant manner.

## 1.3 Structure of the dissertation

I first discuss the relevant stages of vision, limits and models of spatio-temporal vision, and applications of these in computer graphics in Chapters 2 and 3. Chapter 4 describes the behaviour and models of current display technologies. The rest of the dissertation can be divided into two parts. In the first part (Chapter 5), I demonstrate how simple

insights into the limitations of spatio-temporal vision can aid the intelligent design of a high-refresh-rate rendering algorithm (Temporal Resolution Multiplexing). In the second part, I focus on visual modelling, mapping the problem of motion quality to multi-scale models (Chapters 6, 7, and 8), describing how such models can be used to design adaptive rendering algorithms that maximise the perceived quality under a restricted rendering budget.

## 1.4 Publications, presentations, awards

### Publications used in this dissertation

- (1) **Gyorgy Denes**, Akshay Jindal, Aliaksei Mikhailiuk, Rafał K. Mantiuk, “A perceptual model of motion quality for rendering with adaptive refresh-rate and resolution”, *ACM Transactions on Graphics (Proc. of SIGGRAPH 2020)*, 39(4), 133, 2020  
*Paper contribution:* Designed a comprehensive perceptual model of motion quality taking motion blur and judder into account.
- (2) **Gyorgy Denes** and Rafał K. Mantiuk, “Predicting visible flicker in temporally changing images”, *Proceedings of Human Vision and Electronic Imaging conference, 2020*  
*Paper contribution:* Collected psychophysical data and created a novel multi-scale visual model for predicting flicker – a temporal artefact – in images rendered with temporally unstable algorithms.
- (3) **Gyorgy Denes**, Kuba Maruszczyk, George Ash, Rafał K. Mantiuk, “Temporal Resolution Multiplexing: Exploiting the limitations of spatio-temporal vision for more efficient VR rendering”, *IEEE Transactions on Visualization and Computer Graphics*, 2019  
*Paper contribution:* Developed, calibrated and psychophysically validated a novel temporal multiplexing algorithm for reducing the rendering cost and data transfer requirements of high-refresh rate rendering pipelines such as VR.
- (4) **Gyorgy Denes**, George Ash, Huameng Fang, Rafał K. Mantiuk, “A visual model for predicting chromatic banding artefacts”, *Proceedings of Human Vision and Electronic Imaging conference, 2019*  
*Paper contribution:* Designed a multi-scale model for detecting chromatic banding (false contouring) artefacts.
- (5) **Gyorgy Denes**, Kuba Maruszczyk, Rafał K. Mantiuk, “Exploiting the limitations of spatio-temporal vision for more efficient VR rendering”, *ACM SIGGRAPH 2018 Posters (Article No. 21)*

## Publications outside the scope of this dissertation

- (1) Gabriel Eilertsen, Joel Kronander, **Gyorgy Denes**, Rafał K. Mantiuk, Jonas Unger, “HDR image reconstruction from a single exposure using deep CNNs”, *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia 2017)*, 36(6), 2017  
*Paper contribution:* Evaluated CNN-based inverse tone mapping operator (iTMO) designed by collaborators to recover high-dynamic-range images from single-exposure photographs. Evaluation consisted of running and analysing the results of pairwise comparison experiments measuring subjective preference of proposed technique over reference and state-of-the-art iTMOs.
- (2) Kuba Maruszczyk, **Gyorgy Denes**, Rafał K. Mantiuk, “Improving Quality of Anti-Aliasing in Virtual Reality”, *The Computer Graphics & Visual Computing Conference of 2018*

In this dissertation, I do *not* present results obtained or code written by my collaborators (Kuba Maruszczyk, George Ash, Aliaksei Mikhailiuk, Akshay Jindal and Rafał Mantiuk), but acknowledge that regular meetings and feedback from my colleagues as well as my supervisor, Rafał Mantiuk, were invaluable. Also, the motion quality experiments described in Chapter 8 would not have been possible without Aliaksei Mikhailiuk’s active sampling code.

## Presentations

- (1) Predicting visible flicker in temporally changing images  
Rainbow Group Seminar, Cambridge, 2020  
Human Vision and Electronic Imaging Conference (HVEI), San Francisco, 2020
- (2) Temporal Resolution Multiplexing: Exploiting the limitations of spatio-temporal vision for more efficient VR rendering  
Rainbow Group Seminar, Cambridge, 2019  
IEEE VR, Osaka, 2019
- (3) A visual model for predicting chromatic banding artifacts  
Rainbow Group Seminar, Cambridge, 2019  
Human Vision and Electronic Imaging Conference (HVEI), San Francisco, 2019
- (4) Multiplexing schemes for high-refresh-rate and VR displays  
Research students’ presentation (mini conference), Cambridge, 2018



- (5) Introduction to Computer Graphics  
BUPT (Beijing University of Posts and Telecommunications), Summer school, Cambridge 2019

## Patents

- (1) Temporal Resolution Multiplexing Display Systems (7855947-1-PMARTIN)

## Awards

- (1) **Best paper award** at 2020 Human Vision and Electronic Imaging (HVEI) (paper title: *Predicting visible flicker in temporally changing images*)
- (2) **Best journal paper award** at IEEE VR 2019 (paper title: *Temporal Resolution Multiplexing: Exploiting the limitations of spatio-temporal vision for more efficient VR rendering*)
- (3) **Best paper award** at 2019 Human Vision and Electronic Imaging (HVEI) (paper title: *A visual model for chromatic banding artifacts*)
- (4) **Best poster award** at CGVC (poster title: *Improving Quality of Anti-Aliasing in Virtual Reality*)



---

## CHAPTER 2

---

# BACKGROUND ON SPATIO-TEMPORAL PERCEPTION

*“This story I tell you, to let you understand, that; in the observation related by Mr. Boyle, the man’s fancy probably concurred with the impression made by the sun’s light to produce that phantasm of the sun which he constantly saw in bright objects. And so your question about the cause of phantasm involves another about the power of fancy, which I must confess is too hard a knot for me to untie.”*

*Sir Isaac Newton  
Letter to John Locke*

The aim of this dissertation is to demonstrate that insights and models of spatio-temporal vision can be utilised to devise novel rendering algorithms. As such, some understanding of how human vision works is unavoidable. Whilst this topic is rarely considered in the graphics literature, there is substantial knowledge we can rely on in the fields of psychophysics and visual science.

As this dissertation is primarily concerned with high-refresh-rate devices, I focus on spatio-temporal sensitivity and motion perception. Related applications and models of perception are discussed in detail in Chapter 3. First, I provide an overview of the visual system, then describe the family of contrast sensitivity functions, and conclude by separately addressing the four types of artefacts impacting motion quality perception.

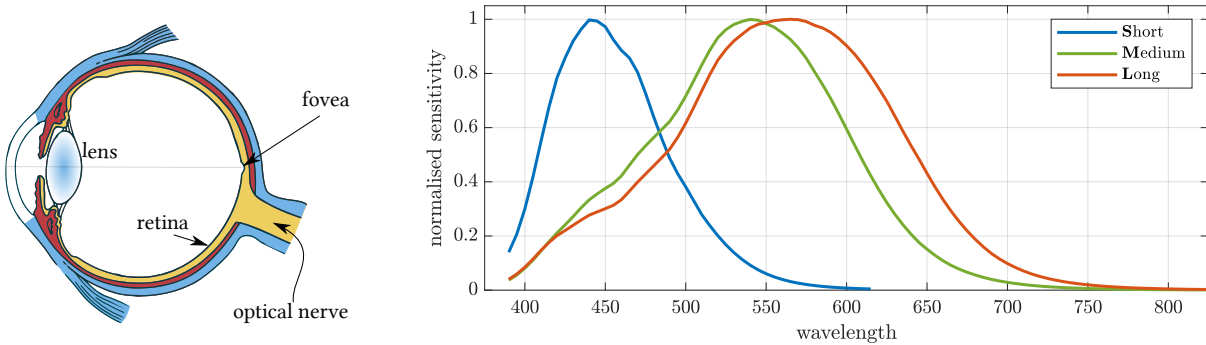


Figure 2.1: Left: biology of the eye. Right: Cone cells’ spectral sensitivity (as a function of light wavelength). Source: Stockman and Sharpe [2000]

## 2.1 Overview of human visual system

The amount of information constantly reaching the human visual system (HVS) with an infinitely detailed world around us is incredible, and in its completeness, unmanageable even to our relatively complex brains. Starting from the retina up to the cortical representation, we find that sensitivity and bandwidth limitations determine our performance at a range of visual functions, be they seemingly trivial or complex such as face recognition, depth perception or colour constancy. Models attempting to capture the behaviour of individual processing stages are mostly based on neurological observations, while end-to-end perception of different stimuli are studied using psychophysical experiments. For a broad yet concise review of the vast HVS literature, I recommend the reader to consult Wandell’s *Foundation of vision* [1995]. In the following sections, I discuss the stages of vision starting from the retina, with emphasis on aspects relevant to spatio-temporal and motion perception.

## 2.2 Early vision

In the first stage of vision, also referred to as *early vision*, information is captured and pre-processed, in a manner comparable to a digital camera [Watson 1990, p.61-74]. As this is the entry point of the pipeline, understanding the limitations of early vision helps to understand the information available to later stages in the visual cortex, where the information is further processed and interpreted by the brain.

Capturing happens on the *retina*, the photosensitive layer at the back of the eye, consisting of an inhomogeneous collection of two types of photoreceptors: *rods* and *cones*. Rods are mainly responsible for vision in low-light conditions (scotopic vision;  $< 0.001 \text{ cd/m}^2$  [Zele and Cao 2015]). These cells provide limited spatial resolution [Wandell 1995, p.46], and do not distinguish between different wavelengths of light; hence, scotopic vision is strictly monochromatic. Rods are primarily found in the peripheral (non-

central) part of the visual field, resulting in higher sensitivity to dim objects when viewed slightly off-centre – a phenomenon utilised by astronomers when observing faint stars. On the other hand, cones operate in well-lit environments (photopic vision;  $> 3 \text{ cd/m}^2$  [Zele and Cao 2015]), providing significantly higher spatial acuity, and are mainly found in the central region (fovea; Figure 2.1-left). The intermediate luminance range between scotopic and photopic vision, when cones have just enough photons to gradually activate and rods are not yet saturated, is referred to as mesopic vision ( $0.001 - 3 \text{ cd/m}^2$  [Zele and Cao 2015]). The luminance range of average content rendered on desktop displays and VR headsets falls mostly within the photopic range.

Humans typically have three types of cones, responding to different wavelengths of light (**Short**, **Medium**, and **Long**). The distinct response curves of cones (Figure 2.1-right) allow the visual system to differentiate between colours, with lacking cones or overlapping response curves causing colour vision deficiency (colour blindness). One limitation of early vision is that the reduction from a domain of continuous wavelength to three response values is not invertible. Consequently, multiple wavelength profiles can correspond to the exact same perceived colour, a phenomenon known as *metamerism*. A well-known example of how insights to the visual system can aid display design is how monitors use three colour primaries (**Red**, **Green**, and **Blue**), relying on metamerism to reproduce a perceptually continuous range of colours.

In the central  $2^\circ$  of the visual field (foveal region), where the cones are most tightly packed, there is an estimated *one cone in every 28"* (arc seconds) [Curcio et al. 1990], yielding approximately 120 cones per visual degree. Assuming a uniform grid, we can compute the maximum resolvable spatial frequency with the Nyquist limit: *i.e.* 60 full cycles of a sine wave over a visual degree, commonly written as 60 cycles per visual degree (cpd). However, the cone mosaic does not follow a uniformly-spaced grid, and eye motion also provides additional information; hence the question of maximum resolution is more complex. For instance, Vernier acuity demonstrates that the HVS can incorporate eye motion and cortical pooling to detect misalignments between elements of the stimulus even when the misalignment is smaller than a single photoreceptor [Soraci and Murata-Soraci 2003, p.51]. It is hence more usual to argue about the visual system’s spatial (and temporal) integration based on end-to-end models of vision.

## 2.3 Visual pathways

Visual pathways (also called visual streams) consist of a series of cells and synapses that carry information from the retina through the optic nerve and the LGN to the visual cortex (V1, V2, V3, V4) [Squire et al. 2009]. The literature identifies several parallel pathways that perform distinct tasks. The most well-known examples are the dorsal and

ventral pathways which are responsible for motion and form perception respectively, also referred to as the “where” and “what” pathways. [Briggs 2017]. Other examples include the Magnocellular and Parvocellular pathways: the Magnocellular pathway is considered insensitive to colour, sensitive to lower spatial *and* higher temporal frequencies. On the other hand, the Parvocellular pathway is sensitive to colour and high spatial frequencies, but has reduced sensitivity to high temporal change [Liu et al. 2006]. The ventral pathway gets its main input from the parvocellular cells, but such visual pathways are often interconnected. Other pathways have been shown to selectively respond to certain scales of size only. The final visual perception is influenced by all the visual pathways – a phenomenon that is challenging to model. A common approach is to focus on measuring the performance at individual tasks in end-to-end psychophysical experiments.

## 2.4 Contrast sensitivity

Overall sensitivity to different stimuli can be described with the family of contrast sensitivity functions. Specifically, sensitivity to an achromatic sine grating as a function of spatial frequency is described by the spatial contrast sensitivity function (CSF). Consistently with previous literature, I will use the CSF to refer to the spatial contrast sensitivity, without qualifying the spatial aspect, unless otherwise specified.

### 2.4.1 Contrast definition

Formally, the most common contrast definition is Weber’s contrast: the *ratio* of a luminance change compared to a baseline luminance level.

$$C = \frac{\Delta L}{L}, \quad (2.1)$$

where luminance is measured in  $\text{cd}/\text{m}^2$ . We use ratios of luminance values rather than absolute luminance differences according to the non-linearity of luminance perception, described by Weber’s Law [Laming 2012, p.177-190].

The smallest change over a background luminance detectable by a human observer ( $\Delta L_t$ ) can be used to describe the threshold contrast ( $C_t$ ). Contrast sensitivity is defined as the reciprocal of the threshold contrast [Wandell 1995, p.135]:

$$S = \frac{1}{C_t} = \frac{L}{\Delta L_t} \quad (2.2)$$

For periodic stimuli, the Michelson contrast definition is used [Kukkonen et al. 1993]:

$$C = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}, \quad (2.3)$$

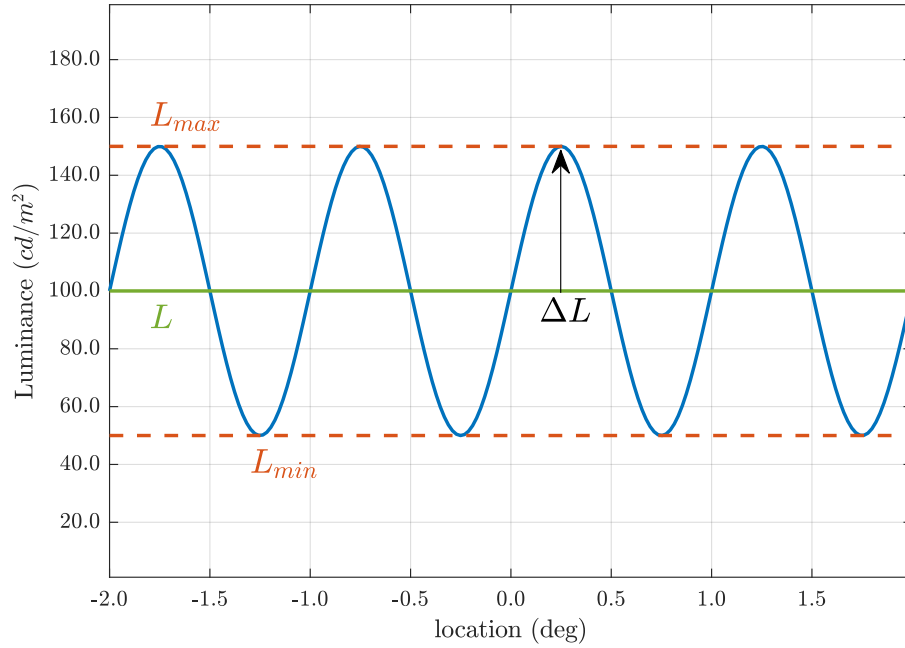


Figure 2.2: Michelson contrast for a sine grating. Taking the mean luminance as  $L$ , and the signal amplitude as  $\Delta L$ , this is equivalent to Weber’s contrast definition.

where  $L_{\max}$  and  $L_{\min}$  are the maximum and the minimum luminance in the periodic signal respectively. For sine gratings, the two definitions are comparable (see Figure 2.2).

## 2.4.2 Spatial contrast sensitivity function

CSF measurements for achromatic sinusoid stimuli show that sensitivity rises from 0 cycles per degree (cpd), peaking on low-to-medium frequencies (3–8 cpd), then falls off exponentially for higher frequencies [Van Nes et al. 1967]. Note, that we are still making a strong distinction between threshold and supra-threshold stimuli; this bandpass-like shape of the CSF is only valid for objects that are around the detection threshold; hence the CSF should not be thought of as a modulation transfer function.

A common interpretation of the complex shape of the CSF is modelling the curve as a combination of responses to multiple spatial frequencies (see Figure 2.3 for visualisation), corresponding to neurons encoding different scales [Wandell 1995, p.136]. Such multi-scale models are inspired by theories that information from early vision is propagated down multiple pathways as discussed in Section 2.3.

The exact shape of the CSF depends on background luminance, orientation, stimulus size, and eccentricity. Some authors measured a reduced sensitivity in higher eccentricities [Thibos et al. 1996; Kelly 1984], verifying the intuition that our photopic vision is sharpest in the central foveal region, degrading rapidly towards the periphery. Virsu and Rovamo [Virsu and Rovamo 1979] argue that the contrast threshold increases linearly with eccentricity, as predicted by the cortical magnification factor (CMF). Extensive studies

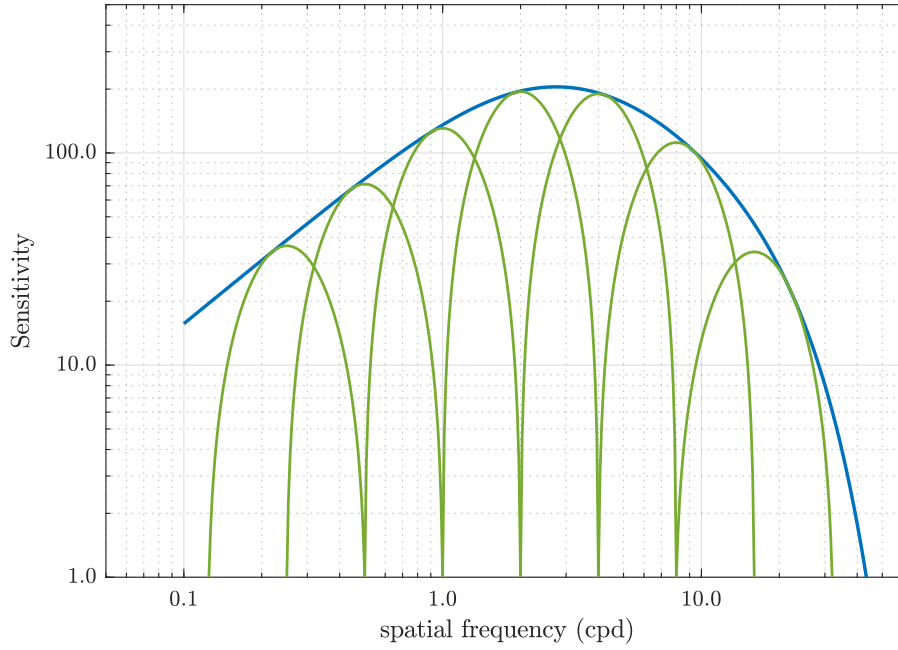


Figure 2.3: Contrast sensitivity function based on Barten et al. [2004]. The overall shape can be interpreted as a non-linear combination of different frequency bands, each peaking an octave apart.

were also conducted to determine sensitivity for colour vision [Kelly 1974; Mullen 1985] and varying background luminance [Kim et al. 2013; Wuerger et al. 2019].

Although there is no comprehensive model of the CSF, some achromatic CSF models exist for standard observers [Barten 2004] that are valid in the fovea. Note, that in practice, these models still require further calibration when being applied on complex stimuli.

### 2.4.3 Temporal contrast sensitivity

For non-static stimuli, sensitivity is described by the temporal contrast sensitivity function (tCSF). de Lange was one of the first people to measure the tCSF as a function of temporal frequencies, by flickering a 2-visual-degree test field over a  $60^\circ$  background [De Lange Dzn 1952; de Lange Dzn 1958]. He found that the shape of the tCSF resembles the shape of the spatial CSF, a band-pass filter with a peak sensitivity at 5-10 Hz. The exact shape depends on a number of factors including background luminance, stimulus size and eccentricity. Crucially, the spatial and temporal dimensions are not separable [Robson 1966; Daly 1998], which can be partially attributed to the fact that both dimensions are affected by the parvocellular and magnocellular pathways in a complex manner. The joint spatio-temporal contrast sensitivity function (stCSF) is shown in Figure 2.5-left.

The visibility of moving objects is better predicted by the spatio-velocity contrast sensitivity function (svCSF) [Kelly 1979]. Here, temporal frequency is replaced with



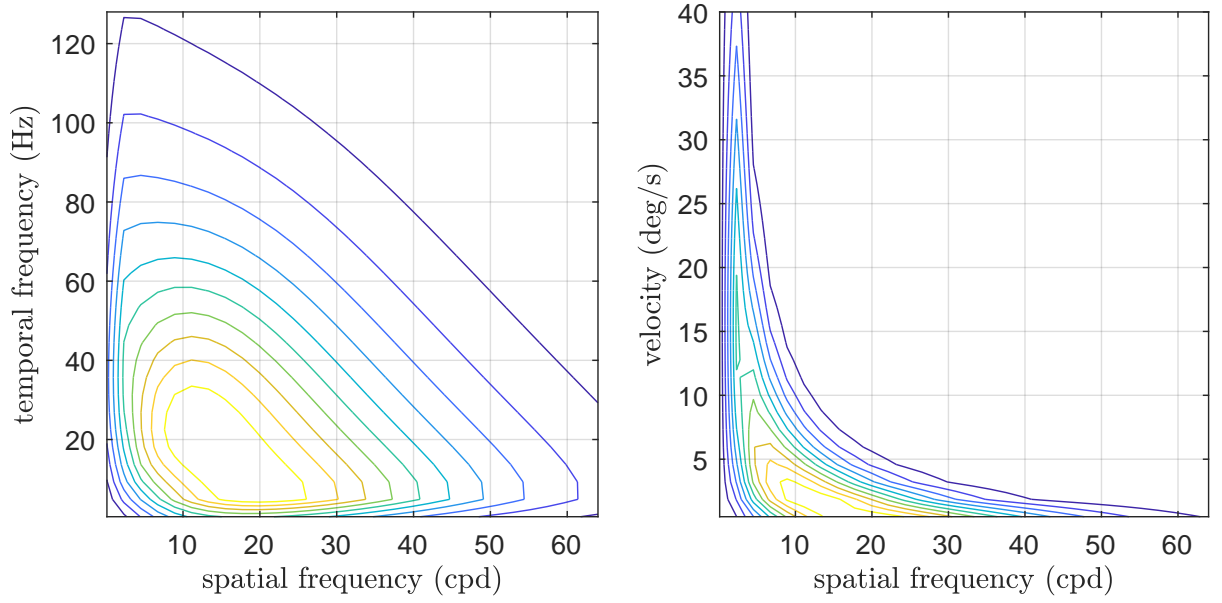


Figure 2.4: Contour plots of spatio-temporal contrast sensitivity (left) and spatio-velocity contrast sensitivity (right). Based on Kelly’s model [1979]. Different line colours represent individual levels of relative sensitivity from low (purple/dark lines) to high (yellow/bright lines).

retinal velocity (in degrees per second). To relate the stCSF, and the svCSF, retinal velocity in the svCSF can be derived as the ratio of the temporal and spatial frequencies in the stCSF. The contour plots of stCSF and svCSF are shown in Figure 2.4. The stCSF plot on the left shows that the contours of equal sensitivity form almost straight lines for high temporal and spatial frequencies, suggesting that the sensitivity can be approximated by a plane. This observation, captured in the window of visibility [Watson et al. 1986; Watson 2013] and the pyramid of visibility [Watson and Ahumada 2017], offer simplified models of spatio-temporal vision, allowing for an insightful analysis of visual system limitations in the Fourier domain.

## 2.5 Temporal integration

Most artificial light sources do not produce a temporally-stable amount of light; they are known to vary with time [Rutherford 2003]. For example, displays with LED light sources control their brightness by switching the source of illumination on and off at a very high frequency, a practice we can refer to as a form of pulse-width-modulation. The perceived brightness of such a flickering display will match the brightness of the steady light that has the same time-average luminance — a phenomenon known as the Talbot-Plateau law [1834]. The Talbot-Plateau law only holds when every frequency of the temporal change is fast enough to be imperceivable. The minimum frequency at which a light is perfectly fused and is perceived as steady is the flicker fusion threshold, or as also known: the

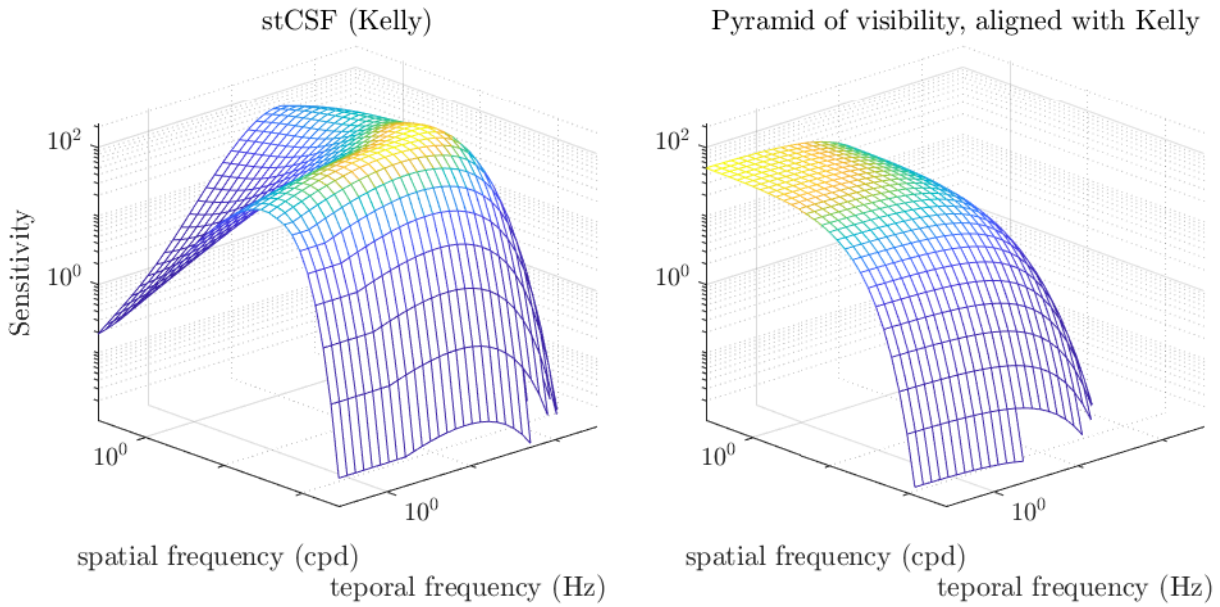


Figure 2.5: Spatio-temporal contrast sensitivity function in the fovea. With increasing spatial and temporal frequencies sensitivity first increases, then falls rapidly. Absolute sensitivity needs to be adjusted with background luminance. Left figure is based on various datasets and fitted models from Kelly and Daly [Kelly 1984, 1964, 1959, 1983; Daly 1998]. Right: the fitted *pyramid of visibility* model, where the approximation holds from medium to high frequencies both in spatial and temporal dimensions [Watson and Ahumada 2017].

critical flicker frequency (CFF). Based on observations of the tCSF it is unsurprising that the CFF depends on multiple factors: it is known to increase proportionally with the log-luminance of a stimulus (Ferry-Porter law), increase with the size of the flickering stimulus, and to be more visible in the *parafovea* (the region between 5-30° from the fovea) [Hartmann et al. 1979].

## 2.6 Motion perception

In order to describe motion perception, I first discuss the relevant types of eye motion, the concept of apparent motion, and conclude by describing the four types of motion artefacts that differentiate perceived motion of objects displayed on a computer monitor from perceived motion in the real world. I will refer to the latter as perfect motion.

### 2.6.1 Eye motion

As described in Section 2.2, the first stage of image capture in the visual system is the retina. However, to explain the retinal image during motion, eye motion also needs to be considered. Eye movements can be divided into five broad categories: fixation,

smooth pursuit eye motion (SPEM), saccades, vestibulo-ocular and optokinetic reflexes, and vergence. For a detailed discussion I refer the reader to consult Leigh and Zee [2015].

## Fixation

Fixation keeps a stationary object aligned with the high-acuity fovea. However, even during the fixation the eyes are not perfectly still. Drifts, tremors and microsaccades with velocities  $0.8\text{--}0.15\text{ deg/s}$  result in small displacements, ensuring that vision does not fade during fixation [Robinson 1964; Martinez-Conde et al. 2004].

## SPEM

When observing a moving object with speeds from  $0.15\text{ deg/s}$  up to  $80\text{ deg/s}$ , the eye follows the object, thus stabilising the image of the object in the foveal region of the retina. This tracking is known as SPEM [Robinson 1965]. SPEM has an approx. 150ms latency, and is imperfect. Inconsistencies between eye motion and the motion of the tracked object result in an unstable retinal location. This in turn, when integrated by the eye under the Talbot-Plateau law, results in retinal blur. The exact amount of such blur depends on the velocity of the tracked object and the nature of its motion. For instance, perfect zero-latency tracking can be achieved for sinusoid, parabolic, and cubic motion trajectories after some learning [Terry Bahill and McDonald 1983] due to the predictive mechanism of the visual system anticipating location and speed [Stark et al. 1962]. However, real-world targets, such as leaves falling or insects flying, often do not follow such predictable patterns. As a fallback, SPEM gets regulated by a feedback mechanism [Lisberger 2010]. Specifically, Lisberger et al. [1981] argues that SPEM mechanism can be viewed as a servomechanism which takes the retinal velocity of the target object into account (retinal error velocity, REV) to induce eye acceleration in order to keep eye velocity close to target velocity. Numerous factors have been demonstrated to affect the accuracy of this feedback mechanism, including traumatic brain injury [Suh et al. 2006] and whether subjects attempt to track the same object with their hands as well [Niehorster et al. 2015].

## Saccades

Saccades are rapid eye movements that shift the line of sight. The larger the saccade, the greater the top speed, reaching  $900\text{ deg/s}$  and up to 100ms [Leigh and Zee 2015]. To prevent a sense of false motion during such high velocities, the visual system ignores much of the intra-saccadic signal, a phenomenon known as saccadic suppression [Castet 2009]. However, studies suggest that the visual system is not fully blind during saccadic motion, and can still perceive temporal change [Davis et al. 2015].

## Vestibulo-ocular and optokinetic reflexes

Eye movements occur during head movements to keep the line of sight on the object of interest. Such reflexes can be further categorised depending on whether the motion is induced by the vestibular system (vestibulo-ocular reflex, VOR), or is based on visual cues (optokinetic reflex)

## Vergence

Frontal vision requires a unique binocular coordination, where the eyes often move in opposite directions. Such vergence movements are primarily generated by disparity between the locations of the a single target on the two retinas. There is also a strong coupling with accommodation (the focal length of the eye) [Leigh and Zee 2015]. A conflict between vergence and accommodation cues can result in reduced depth perception, an issue present in most 3D displays.

### 2.6.2 Apparent motion

Computer monitors, much like traditional TV sets, display a sequence of static frames to the viewer, yet a human observer perceives continuous motion. This *apparent motion* has been studied within a number of fields, with computer graphics primarily focusing on the questions of frame duration and refresh rate.

Following signal-processing terminology, the reduction of a continuous signal in time (perfect motion) to a discrete set of frames can be considered sampling. The action of the sequence of static images in a continuous time domain can be modelled as a reconstruction problem, where display technologies differ in the shape of the reconstruction window. LCD monitors approximately show and *hold* a static frame until the next frame becomes available, effectively reconstructing with a rectangle-like window. In contrast, the reconstruction window of CRT monitors was determined by the phosphor layer's time response – short burst with a long tail. Stroboscopic displays only flash the stimulus for a brief  $\Delta t$  time, much shorter than the whole frame duration. Intensity of the signal then needs to be enhanced by a factor of  $1/\Delta t$  as stipulated by the Talbot-Plateau law.

Watson et al. modelled *apparent motion* with their proposed window of visibility [Watson et al. 1986], a diamond-shape in the spatio-temporal domain bound by the critical flicker frequency and the eye's maximal spatial resolution in the temporal and spatial domains respectively. Watson observed that sampling introduces periodic repetition in the spatio-temporal frequency spectrum, and proposed that motion is only perceived smooth when the spectrum of the up-sampled signal contains no copies inside the window of visibility (see Figure 2.7). The duty-cycle of the display – *i.e.* the fraction of the period when the display is active within the frame time, has no observable impact here.

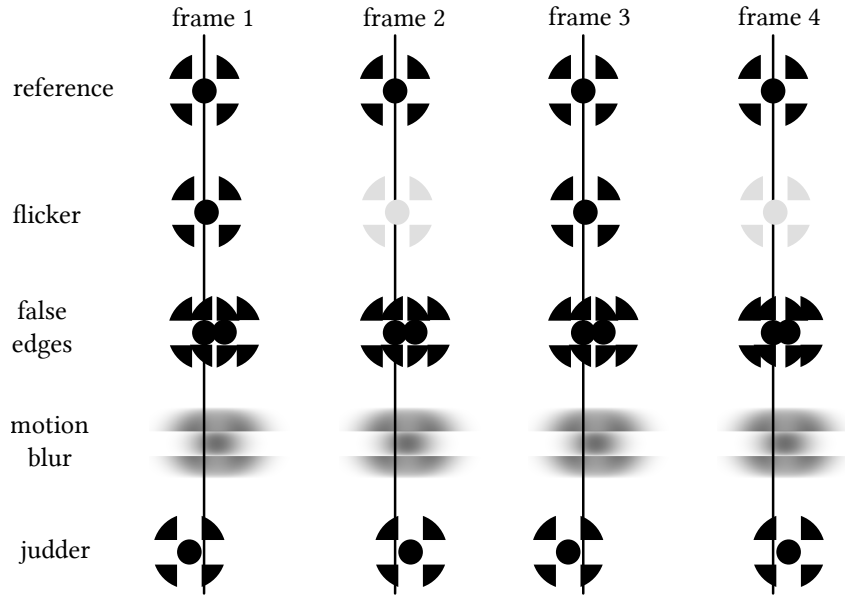


Figure 2.6: Motion artefacts (rows 2-5) compared to perfect motion (top row) for consecutive frames (columns). Flicker: luminance change in consecutive frames; false edges: multiple copies of the original object; motion blur: loss of high-frequency detail; judder: object location is inconsistent in consecutive frames (vertical lines indicate reference locations for each frame).

### 2.6.3 Motion artefacts

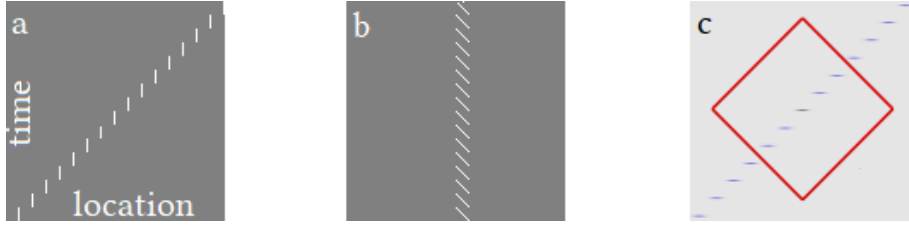
Although the human brain accepts traditional cinema films to have apparent motion, even an untrained observer can notice a difference when compared to motion quality in the real world (perfect motion). The differences between apparent motion and perfect motion can be divided into four categories: (1) flicker, (2) false multiple edges (ghosting), (3) motion blur, and (4) non-smooth or juddery motion (also known as strobing or stutter) [Daly et al. 2015]. In the following paragraphs I discuss each of these in turn. Please see Figure 2.6 for a visualisation.

#### 2.6.3.1 Flicker

Flicker artefacts arise when the luminance of a stimulus changes periodically at a temporal frequency and contrast objectionable by the human eye. The visibility of flicker depends on a number of factors, but is generally well-captured by the spatio-temporal contrast sensitivity function.

#### 2.6.3.2 False edges

Even with the rate of change above the CFF, artefacts might manifest as false edges; *i.e.* the integrated image on the retina containing multiple copies of the original stimulus. Such ghosting artefacts can occur if a low-persistence displays repeats the same frame (at



Computing retinal signal of a moving line in the Fourier domain.

- (a) sampled location in screen coordinates,
- (b) retinal signal — signal from (a) accounting for perfect SPEM
- (c) Fourier transform of (b).

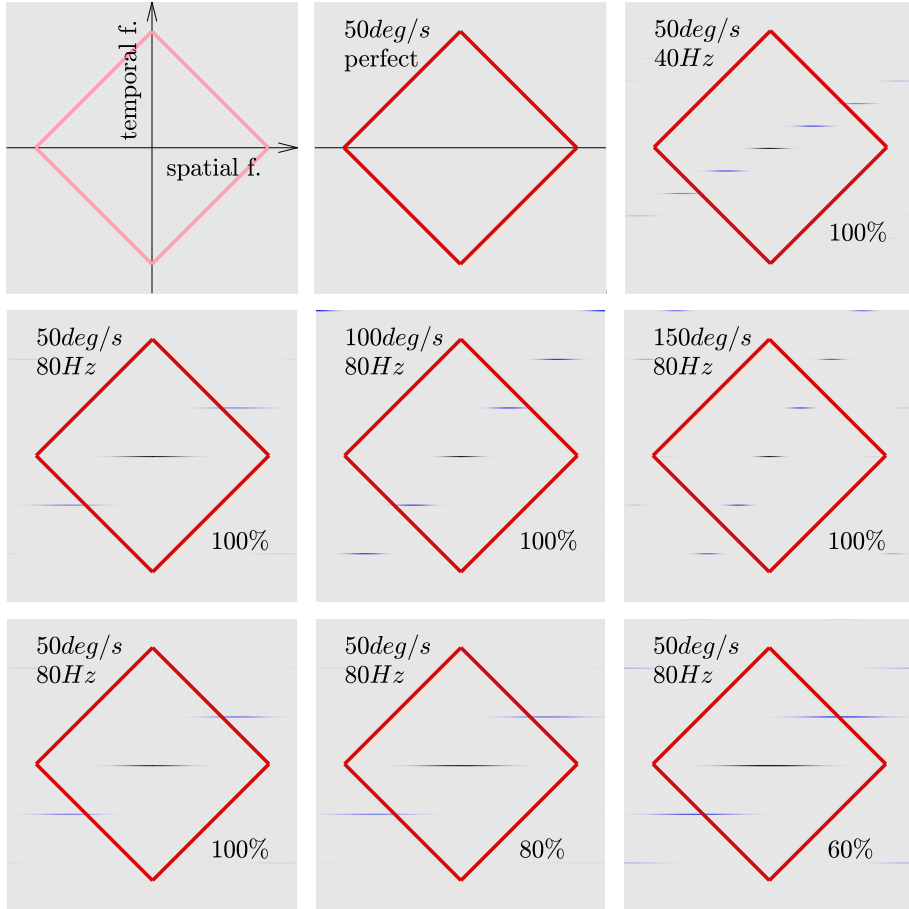
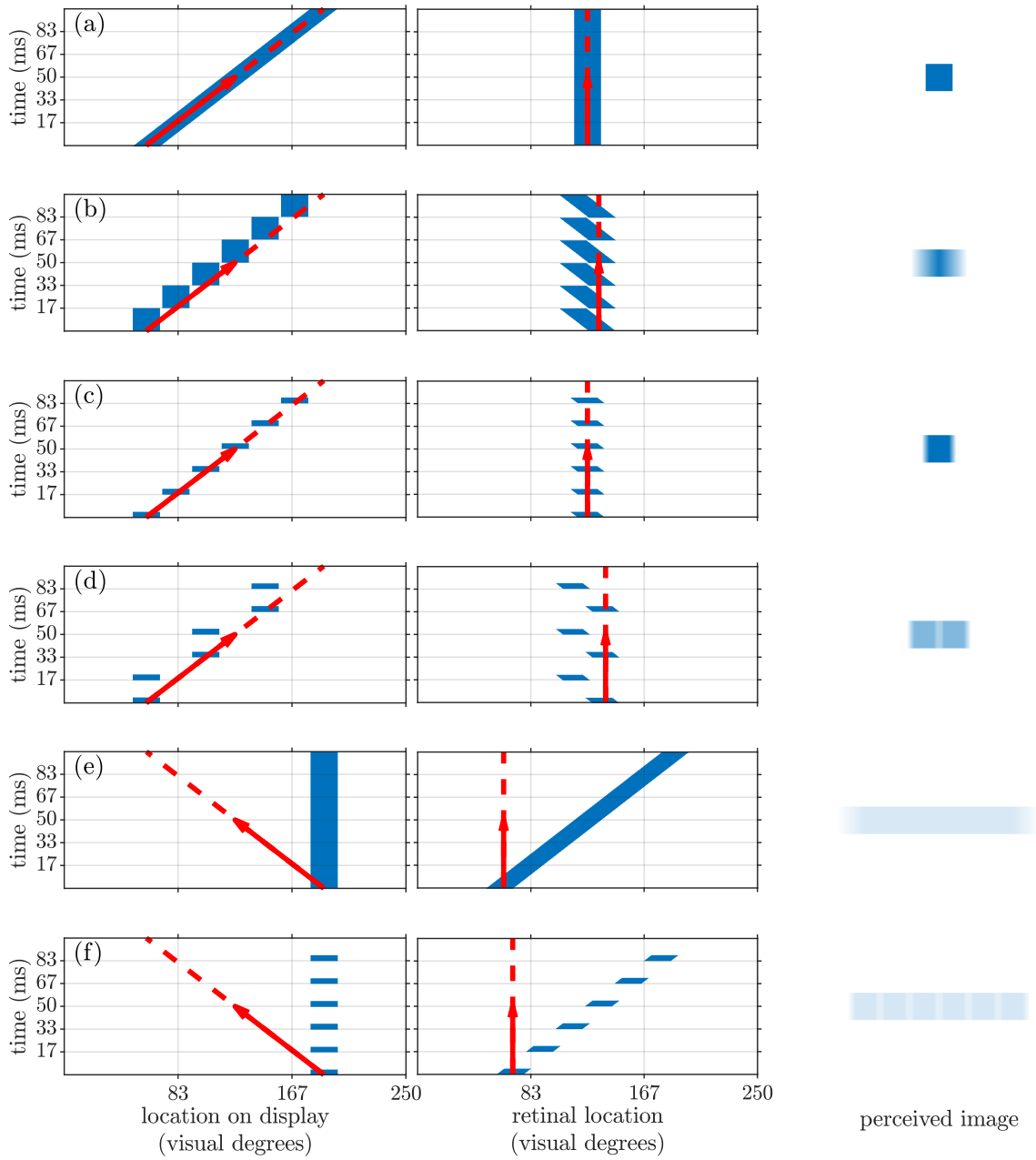


Figure 2.7: Spatio-temporal frequency analysis of a the retinal image when SPEM follow-ing a 1-pixel-wide line moving left to right with constant velocity, sampled by a display. Top subplot shows the derivation with the retinal signal reduced to one spatial and the temporal dimension, transformed to the Fourier domain. Due to discrete sampling, aliasing copies of (blue) of the spectrum of the original signal (black) are visible. Analysis illustrates when sampled on different refresh rates (top row), moving with different velocities (middle row), and with low-persistence displays of varying duty cycle (bottom row). Labels indicate object velocity in degrees per second, refresh rate in Hz and duty cycle in percentage. Red diamonds show the window of visibility [Watson et al. 1986], i.e., the threshold frequency in the spatio-temporal domain outside of which the visual system cannot detect artefacts.



least) twice. A similar artefact can be observed with DLP projectors, where the colour wheel presents the red, green and blue images subsequently. Alternatively, phantom arrays (multiple copies of the same object) can be perceived during saccades at frame refresh rates far beyond the CFF (500 Hz-1kHz) [Davis et al. 2015; Roberts and Wilkins 2013a]. For a visualisation, see Figure 2.8.

### 2.6.3.3 Blur

Most blur artefacts can be attributed to LCD displays holding an image for the full duration of a frame. When the gaze follows a moving object, it moves continuously over pixels that are stationary over a duration of a single frame. At higher refresh rates, this is perceived as blur rather than non-smooth motion because the visual system integrates the image over time [Tourancheau et al. 2009]. The amount of blur increases proportionately with velocity, and decreases with refresh rate. Hence, on VR headsets, where high velocities are common with head motion, motion blur is more prominent. This results in simulation sickness [Anthes et al. 2016]. Stroboscopic low-persistence displays perform much better in terms of motion sharpness than hold-type (high-persistence) displays (Figure 2.8). This reveals a recent example of how perceptual insights can contribute to display design. VR headsets are one of the earliest adaptors of OLED displays due to their low-persistent behaviour.

The other source of blur is eye motion. Predictable targets can be followed quite accurately as described in Section 2.6.1. However, real-world movement is often unpredictable, resulting in imperfect SPEM tracking and the target object consequently blurring on the retina.

Further blur artefacts due to imperfections in the optics of the eye are not discussed here, as they do not vary as a function of motion.

Watson and Ahumada [2011] provides an excellent review of studies investigating the visibility of blur. With a focus on Gaussian blur (described by the standard deviation), the authors unify available psychophysical experimental data, and discuss proposed models for the visibility of blur. In Chapter 8, I follow similar principles to their visual contrast energy (ViCE) model by computing energy after modulating the blurred signal with the contrast sensitivity function. However, the field I focus on in this dissertation – high-refresh-rate rendering – requires a more careful consideration of eye blur as a function of motion velocity and SPEM predictability.

### 2.6.3.4 Judder

At low refresh rates the illusion of motion breaks, and individual frames become visible. This creates artefacts known as judder, stutter or strobing motion. Judder is caused by the discrete temporal samples of the display (frames), which produce aliasing; *i.e.* copies



of the spectrum of the original signal will appear in the frequency domain (see Figure 2.7). The temporal frequency of these aliasing copies depends on the refresh rate of rendering, whereas the spatial frequency is determined by motion velocity. Therefore, judder can be observed when motion velocity is high or when refresh rate is low.

## 2.6.4 Motion sharpening

Even with predictable SPEM tracking, the eye lags behind the target object. Daly et al. [1998] estimates the ratio of target velocity and eye velocity as 0.82 (also known as the average velocity gain). The retinal blur resulting from the velocity discrepancy would intuitively imply reduced contrast sensitivity for moving objects. However, no drop in sensitivity was observed for velocities up to  $7.5 \text{ deg/s}$  [Laird et al. 2006] and only a moderate drop of perceived sharpness was reported for velocities up to  $35 \text{ deg/s}$  [Westerink and Teunissen 1990]. Blurred images appeared sharper when moving with speeds above  $6 \text{ deg/s}$  and the perceived sharpness of blurred images was close to that of sharp moving images for velocities above  $35 \text{ deg/s}$  [Westerink and Teunissen 1990]. This effect, known as *motion sharpening*, can aid us to see sharp objects when retinal images are blurry because of imperfect SPEM tracking. Takeuchi and De Valois demonstrated that this effect corresponds to the increase of luminance contrast in medium and high spatial frequencies in moving objects [Takeuchi and De Valois 2005]. They also demonstrated that interleaved blurry and original frames can appear close to the original frames as long as the cut-off frequency of the low-pass filter is sufficiently large. Such examples once again highlight the complexity of the visual system.

## 2.7 Summary

Successful computer graphics algorithms need to take into account the eventual consumer: the human eye. Understanding some of the individual components of the visual system has been shown to benefit display design. For example, the design of display primaries that produce a range of metameric colours, or low-persistence panels, which reduce motion artefacts. However, the human visual system is complex, consisting of a set of stages from early vision (capture and pre-processing) to high-level cortical interpretation. Biological understanding of individual stages such as the behaviour of rods and cones on the retina or gaze tracking measurements provide useful insights, but psychophysical experiments are required to measure and model the end-to-end perception of stimuli. Contrast sensitivity functions for spatially, temporally, or spatio-temporally changing images (CSF, tCSF, stCSF) are particularly well-researched tools for modelling the limitations of the visual system. The other highly relevant framework for analysing motion quality, is separating artefacts into (1) flicker, (2) false multiple edges (3) motion blur, and (4) judder.

In the next chapter, I describe how the principles introduced here have aided the design of existing display technologies and graphics algorithms. In Chapters 5–8 I also rely heavily on the family of contrast sensitivity functions and the classification of motion artefacts to propose novel algorithms and to design visual models for motion.

---

---

## CHAPTER 3

---

### RELATED WORK

*“The whole is other than the sum of the parts”*

*Kurt Koffka*

Understanding the fundamental limitations of early vision and the visual system’s contrast sensitivity provides the essential insights that can be modelled, inverted and exploited. In this chapter I discuss how perceptual knowledge has been applied in the field of computer graphics and display design. I start by outlining attempts to establish the ultimate spatial and temporal monitor resolution, then discuss high-refresh-rate technologies, and examine existing multiplexing algorithms. At the end of the chapter I also explain how perceptual models can be used to predict a perceivable difference between static images.

### 3.1 Critical resolution

Analogously to the critical flicker frequency, knowledge of early vision and psychophysical results can be applied to establish the critical spatial and temporal resolution.

#### 3.1.1 Spatial resolution

The most well-known phrase in search of the critical spatial display resolution (a resolution beyond which no improvement yields a perceived benefit) is one that comes from the industry: Apple’s retina displays. The term was first coined for to the 3.5”  $960 \times 640$  pixel iPhone 4 display. Although the number of pixels is not especially high compared

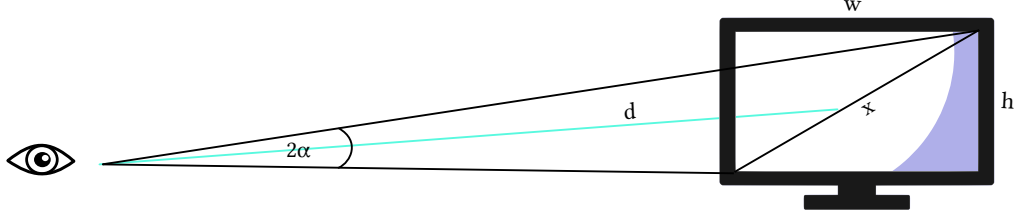


Figure 3.1: Computing the angular resolution with monitor diagonal ( $x$ ) in pixels ( $x_{px}$ ) or cm ( $x_{cm}$ ), and viewing distance  $d$ .

to more current figures; in terms of angular resolution, the iPhone 4 provided over 120 pixels per visual degrees (ppd) from a viewing distance of 53cm. This in turn can show 60 cpd, which coincides with the limit of spatial acuity.

The display resolution  $r$  (in pixels per degrees) can be computed as:

$$r = \frac{x_{px}}{2 \arctan\left(\frac{x_{cm}}{2d}\right)}, \quad (3.1)$$

where  $x_{px}$  and  $x_{cm}$  are the screen diagonal in pixels and cm respectively,  $d$  is the viewing distance, and  $\arctan$  is in degrees. The minimum viewing distance to achieve 120 ppd can be derived as:

$$d = \frac{x_{cm}}{2 \tan\left(\frac{x_{px}}{240^\circ}\right)}, \quad (3.2)$$

In contrast, a standard 15.6" laptop with Full HD ( $1920 \times 1080$ ) display needs to be viewed from 123cm to get the same angular resolution, but a 4K panel of the same size could be reasonably called a retina display with a minimum viewing distance of 60cm. Such numbers indicate that in terms of resolution, desktop and mobile displays are close to the limits of the spatial acuity of the eye. At the same time, latest VR headsets (HTC Vive Pro with  $1440 \times 1600$  pixels per eye,  $110^\circ$  field of view) only have an appalling 20 ppd resolution.

However, it is worth noting that 60cpd is not a hard limit; contrast and luminance can affect contrast sensitivity at such high frequencies; also, the 60cpd estimate ignores the non-uniform structure of the cone mosaic, as well as eye motion. For instance, as highlighted in Section 2.2 during Vernier acuity tasks, humans can detect misalignments between line segments even beyond half an arc minute.

### 3.1.2 Temporal resolution

The temporal domain has historically received less interest than the spatial domain in many walks of research and development. While spatial resolution has been gradually increasing over the years, temporal resolution (refresh rate) has remained mostly fixed at

60 Hz. The renaissance of VR and the increasing popularity of gaming monitors has only recently started expanding this to 90 Hz, then 120 Hz with latest LCD panels operating on 165 Hz and 240 Hz.

Numerous previous attempts have been made to establish the critical monitor refresh rate, beyond which improvements yield no perceptual benefit. Noland et al. [2014] applied traditional sampling theory, combining the CSF with a simple model of eye motion, employing the Nyquist limit to derive the refresh rate that is indistinguishable from perfect motion. Their model for an LCD display predicts that while 140 Hz is sufficient for untracked motion, tracked motion requires at least 700 Hz for the illusion of perfect motion. The authors highlight that the figures should only be considered approximate, partly due to the limitations of Daly’s model of SPEM [1998]. Deeper knowledge of the SPEM mechanism suggests that the nature of motion (predictable vs. unpredictable) is also likely to affect this figure. Kuroki et al. [2006; 2007] arrived at a more conservative estimate using psychophysical measurements showing that at least 250 Hz is required to completely remove motion blur and judder. Such refresh rates are unfortunately still beyond the capabilities of most consumer GPUs and monitors, and the threshold numbers provide little intuition as to how exactly the perceived quality of motion increases above 60 Hz.

The perception of motion quality has been measured in a number of experiments. Notably Mackin et al. [2016] isolated display blur and temporal artefacts such as flicker and judder, collecting mean impairment scores (MIS) for each as a function of object velocity and monitor refresh rate. The authors concluded that for object velocities below  $60 \text{ deg/s}$ , about 50% of the critical refresh rate (at which no artefacts are detected) could provide an *acceptable* MIS, however, they do not provide any guidelines as to how different motion artefacts contribute to the overall perceived quality below such threshold.

In the film industry, refresh rate can be considered an artistic tool. There is evidence for a learned cultural bias in the context of cinemas, sometimes described as the soap opera effect, indicating that observers prefer traditional 24 Hz content over higher refresh rates. Simultaneous management of objectively better motion quality and viewer expectations can be achieved by emulating continuously varying frame rates [Templin et al. 2016]. An acquired taste similar to the soap opera effect has not been reported for computer games or other applications, therefore it is reasonable to assume that discussions and models of motion quality in this dissertation are valid for content outside the film industry. For this reason, I do not consider video quality metrics such as VQM and MOVIE [Seshadrinathan and Bovik 2010], or STR-RED [Soundararajan and Bovik 2012] to be directly applicable to rendered content.

Perceived motion quality of panning (horizontal movement on the screen) was investigated recently by Chapiro et al. [2019]. The authors measured subjective motion

artefact scores for refresh rates typical of modern televisions (30, 60, and 120 Hz) across a range of luminance (2.5–40 cd/m<sup>2</sup>) and camera panning speeds (2–6.6 deg/s). A trivariate quadratic empirical model was shown to fit their data well. In contrast to their study, which considered cinematographic content, my work is focused on computer graphics rendering. The range of refresh rate and velocity values I consider are more representative of computer games (up to 165 Hz with motion speeds up to 45 deg/s) with predictable and unpredictable SPEM. In Chapter 8 I demonstrate how the empirical model of Chapiro et al. is not suitable for such range.

## 3.2 Temporal display technologies

While most content is generated at 60 Hz, display manufacturers have been exploring options to increase the perceived motion quality through blur reduction and adaptive refresh rate technologies.

### 3.2.1 Blur reduction

Blur can be mainly attributed to eye motion over an image that remains static for the full duration of a frame [Feng 2006]. When the eye follows a moving object, the gaze smoothly moves over pixels that do not change over the duration of the frame. This introduces blur in the image that is integrated on the retina, an effect known as *hold-type blur* (refer to Figure 2.8 for the illustration of this effect). Hold-type blur can be reduced by shortening the time pixels are switched on, either by flashing the backlight [Feng 2006], or inserting black frames (BFI). Both solutions, however, reduce the peak luminance of the display and may result in visible flicker.

Virtual reality headsets are bringing back the notion of stroboscopic or low-persistence displays, where the duty cycle of the display is reduced. The display is active only during a fraction of the frame time, but the peak luminance is increased proportionately according to the Talbot-Plateau law. Low-persistence displays reduce motion blur during SPEM and head motion, as discussed in Section 2.6.3.3, which in turn is claimed to reduce simulation sickness [Ishan Goradia et al. 2014]. The refresh rate must be sufficiently high to prevent visible flicker on the given display luminance.

OLED-based mobile phone displays can also provide two modes of operation: a standard mode where each frame is held for the full duration (maximising display brightness), and a VR mode where the display produces a low-persistence behaviour.

Nonlinearity compensated smooth frame insertion (NCSFI) attempts to reduce hold-type motion blur while maintaining peak luminance [Chen et al. 2006]. The core algorithm generates a sharpened and blurred image pair, which is fused into the original image with reduced motion blur. NCSFI is designed for 50–60 Hz TV content and, as demonstrated in

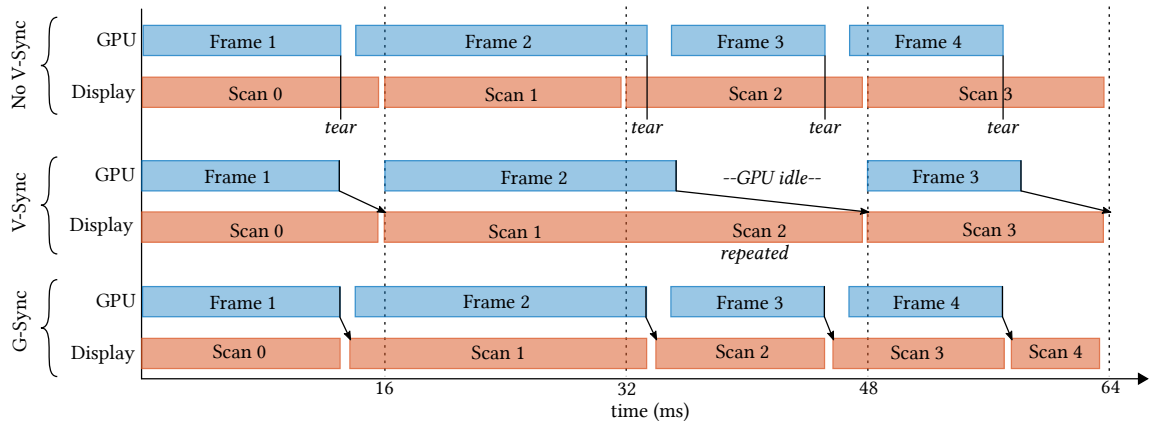


Figure 3.2: Illustration of display sync strategies assuming a screen with approx. 60 Hz refresh rate. No V-sync introduces tearing; V-Sync can result in repeated frames and idle GPU cycles; G-Sync provides the best results by synchronising scanout with frame generation.

Chapter 5, produces ghosting artefacts for high angular velocities typical of user-controlled head motion in VR.

Other attempts have been made in the past to blur in-between frames to improve coding performance [Fujibayashi and Boon 2008b]. These methods rely on the visual illusion of motion sharpening which makes moving objects appear sharper than they physically are. However, no such technique has been incorporated into a coding standard. One issue is that at low velocities motion sharpening is not strong enough, leading to a loss of sharpness.

### 3.2.2 Display sync (V-Sync, G-Sync, Free-Sync)

The general process of image presentation is as follows: (1) the graphics processing unit (GPU) receives draw commands, (2) GPU executes draw commands and stores the result in the frame buffer, (3) monitor updates its content. In practice, the above operations happen concurrently. While the monitor is updating (non-atomic, usually executed row-by-row), the graphical application fires new draw commands for the next frame at the GPU, which then starts working on them in parallel. The order of draw calls is preserved, but draw commands are usually asynchronous.

With a single frame buffer, the monitor might find itself reading out a half-complete frame which can lead to strong flickering artefacts. For this reason, most applications employ double-buffering, where the GPU outputs the in-progress result of draw calls to a back buffer, and only swaps to the front buffer once the frame is complete. If the swap happens more frequently than monitor updates (these occur at 60 Hz on standard desktop displays), or the updates are not synchronised, tearing artefacts might occur. That is, different parts of the monitor display inconsistent versions of the frame buffer.

Synchronising buffer swaps with monitor reads (V-Sync) can effectively prevent tearing [Poth et al. 2018], but might cause the GPU to block while waiting for the update. This can decrease performance and introduce input lag. Even worse, when content generation falls only slightly below 60 Hz, V-Sync will cause the monitor to repeat every other frame, forcing the graphics pipeline to sit idle almost 50% of the time, while introducing judder, blur and potential ghosting in the images. For an illustration of the different strategies, see Figure 3.2.

G-Sync (NVidia) and FreeSync (AMD) are two similar techniques to eliminate both screen tearing and judder. In both technologies, instead of forcing the application to synchronise the rendering pipeline with the monitor updates, the monitor needs to try and synchronise its updates to the frame generation. Refresh rates are no longer fixed, the monitor synchronising circuit should be able to update at any arbitrary refresh rate within the capabilities of its LCD crystals [Poth et al. 2018]. This yields a continuous range of refresh rates from 20 Hz to 165 Hz on current generation hardware (ASUS ROG SWIFT PG279Q). However, even on modern LCD panels, frame updates are not instantaneous. As in G-Sync and FreeSync, the frame rate is completely dictated by the GPU, the monitors need to be able to reasonably predict the presentation time of the next frame to match backlight modulation and overdrive accordingly. When the system is faced with highly irregular and frequently changing frame rates, prediction might fail, and transition artefacts might occur. Furthermore, the prediction system is an isolated unit on the monitor hardware, and there is currently no mechanism for the GPU to aid the prediction mechanism with its own estimated refresh rate.

### 3.3 Multiplexing techniques

One way to overcome the computational and bandwidth limitations of consumer hardware is to assume a fixed computational budget. In such a setup there are a number of factors that contribute to the overall rendering cost including scene complexity, shading and texturing quality, bit-depth, refresh rate, and scene resolution [McCarthy et al. 2004]. Increasing the sampling in one dimension entails reduction in another. Such a problem can be modelled computationally, but the overall perceived quality requires an understanding of relevant perceptual principles. For a comprehensive survey of this matter, I recommend consulting [Masia et al. 2013]. In this section, I review a few notable techniques with primary focus on algorithms that consider the temporal domain.

#### 3.3.1 Resolution vs. colour

Human eye is more sensitive to change in luminance than colour, and chromatic contrast sensitivity drops rapidly with higher spatial frequencies [Wandell 1995]. Images can be





Figure 3.3: Illustration of chroma-subsampling in Yuv colour space. Blurring either or both the chroma channels ( $u$  and  $v$ ) results in minor discolourations, while blurring the luminance channel ( $Y$ ) causes highly objectionable quality loss.

decomposed into a luma (brightness) and two chroma (colour) channels, usually by transforming from red, green, and blue (rgb) to a colour space such as  $YC_bC_r$ ,  $Yuv$  or  $ITP$ . Note that the distinction between luminance and luma originates from the attempt to differentiate between physical luminance and screen-relative brightness respectively.

Resolution reduction in the luma channel is much more visible than resolution reduction in either or both the chroma channels (see Figure 3.3). A wide-spread implementation of this observation in the image and video compression community is chroma-subsampling, found in both the JPEG and MPEG standards. Here, down-sampling is performed in  $YC_bC_r$  both horizontally and vertically usually by a power of two. The downsampling factors are customisable and usually denoted as L:H:V, where L is the horizontal sampling reference, H is the horizontal chroma factor (relative to the first digit), and V is the vertical chroma factor [Poynton 2002]. For instance, 4:4:4 indicates a full-resolution image (both luma and chroma), 4:2:2 denotes full luma and half the chroma resolution both horizontally and vertically. Assuming an equal number of bits in each channel, 4:2:2 yields 50% less data usually without introducing visible artefacts.

A display technique exploiting chroma-subsampling is Samsung’s PenTile™ technology, popularised by the smartphone and VR markets [Elliott et al. 2005]. The authors note that the eye has higher sensitivity to luminance when it comes to high-frequency signals. Therefore, for each blue sub-pixel, they suggest two green and two red sub-pixels. In practice, the number of red sub-pixels is also halved, resulting in a Bayer pattern (Figure 3.4). Resolution is reported in terms of the green sub-pixels, and the value of the red and blue *super-pixels* are shared by the neighbouring pixels.

### 3.3.2 Resolution vs. refresh rate

Claypool et al. investigated the effect of latency and trading off resolution for refresh rate in the context of video games, specifically within the first-person shooter (FPS) genre [Claypool and Claypool 2007, 2009], where players engage in fast-paced virtual combat, viewing the scene through the eyes of the avatar. Large-scale user studies revealed that the refresh rate has a significantly larger influence on task performance compared to res-

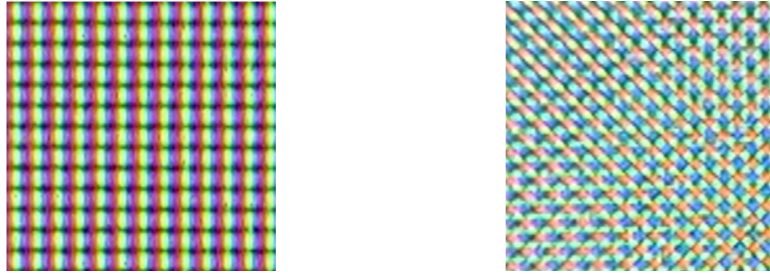


Figure 3.4: Traditional sub-pixel alignment (left) splits each pixel in three equal parts for red, green and blue primaries. PenTile™ pixels have a diamond shape with twice as many green sub-pixels as red or blue.

olution. On frame rates as low as 3-7 Hz users could not adequately target opponents, and there were clear task performance benefits of increasing the refresh rate up to 60 Hz. Perceptual quality and playability gathered with post-experiment questionnaires revealed a similar but less pronounced trend [Claypool and Claypool 2007, 2009]. Unfortunately no measurements were made beyond the capabilities of a standard 60 Hz monitor. Higher refresh rates result in reduced game latency, another factor known to affect task performance in FPS games [Beigbender et al. 2004].

Controlling rendering quality under a constrained budget can be formulated as an optimisation problem, where the free parameter is the ratio of refresh rate and resolution [Debattista et al. 2018]. Preference data collected in a two-alternative-forced-choice (2AFC) design indicated that the optimal ratio is dependent on the computational budget. The authors provide a mathematical model for selecting the optimal refresh rate, but no consideration was given to the underlying mechanisms of the visual system that influences preference. Notably, the experiment established a single refresh rate value over a complete animation clip. As the perceived motion quality is known to change with scene velocity, it is reasonable to expect that the optimal ratio is also a function of that velocity. For example, a fully stationary scene would only benefit from high resolution, whereas fast motion can produce highly-objectionable judder when rendered at low refresh rates. In Chapter 8, I demonstrate that incorporating knowledge of object motion can predict a better trade-off and allows the development of adaptive algorithms that dynamically change the refresh rate of the animation.

### 3.3.3 Temporal multiplexing

Temporal multiplexing, taking advantage of the finite integration time of the visual system and trading off temporal resolution for another rendering dimension, is a widely-researched field.

### 3.3.3.1 Resolution vs. time

In order to increase the perceived spatial resolution, the content on a high-refresh-rate monitor can be rapidly varied. For instance, Didyk et al. [2010a] considers smoothly moving images and optimises up to three consecutive frames under the assumption of perfect temporal integration and smooth pursuit eye motion. On their 120 Hz test monitor this introduced low-contrast flicker at 40 Hz — well below the CFF. Authors mitigate this with a multi-scale CFF predictor, conservatively disabling quality improvements when any probability of flicker is detected. Alternatively, Berthouzoz et al. [2012] show that for stationary content, the display can be *wobulated* (oscillated) to achieve the same results.

### 3.3.3.2 Colour vs. time

Temporal multiplexing is also widely used to *increase perceived bit-depth* via spatio-temporal dithering [Mulligan 1993]. Colour is commonly encoded in 8 bits (256 discrete values) per colour channel. In dithering techniques each pixel is rapidly alternating between two integer values to produce intermediate luminance values, thereby extending the bit-depth to 9 bits. Joint alternation of neighbouring pixels can achieve 10 bits. The luminance change ( $\Delta L$ ) in these schemes is usually small compared to the average luminance ( $L$ ); therefore, the contrast of the spatio-temporal change is also small. The eye’s contrast sensitivity is low on high spatial and temporal frequencies, which makes such schemes applicable even on standard monitors.

Digital Light Processing (DLP) projectors also rely on temporal multiplexing to produce colour images. The core digital micromirror device (DMD) can update whether or not to reflect the white light source thousands of times a second – a combination of dithering and a rotating *colour wheel* is then used to produce 8-bit colour images. The colour wheel and the light source jointly determine the colour gamut of the projector. Recent research attempts to adaptively change the colour gamut to match the displayed content [Kauvar et al. 2015]. This is achieved by altering the spectral power distribution of the light source, keeping the colour wheel intact.

### 3.3.3.3 Temporal coherence

Since consecutive frames in an animation sequence tend to be similar, rendering cost can be reduced by exploiting temporal coherence. A comprehensive review of related techniques can be found in [Scherzer et al. 2012]. Here, we focus on the methods that are the most relevant for VR application: reverse and forward re-projection techniques.

The rendering cost can be significantly reduced if only every  $k$ -th frame is rendered, and in-between frames are generated by transforming the previous frame. *Reverse re-projection* techniques [Nehab et al. 2007] attempt to find a pixel in the previous frame for

each pixel in the current frame. This requires finding a re-projection operator, mapping pixel screen coordinates from the current to the previous frame and then testing whether the current point was visible in the previous frame. Occlusion can be tested by comparing depths for the current and previous frames. *Forward re-projection* techniques map every pixel in the previous frame to a new location in the current frame. Such a scattering operation is not well supported by graphics hardware, making a fast implementation of forward re-projection more difficult. This issue, however, can be avoided by warping the previous frame into the current frame [Didyk et al. 2010b]. This warping involves approximating the motion flow with a coarse mesh grid and then rendering the forward-re-projected mesh grid into a new frame. Since parts of the warped mesh can overlap with the other parts, both spatial position and depth need to be re-projected and the warped frame needs to be rendered with depth testing. Recent techniques employ bidirectional optical flow to warp frames and use them along with contextual information to synthesise interpolated frames with the help of a convolutional neural network [Niklaus and Liu 2018].

Commercial VR rendering systems use re-projection techniques to avoid skipped and repeated frames when the rendering budget is exceeded. These techniques often involve rotational forward re-projection [Vlachos 2015]. Rotational re-projection assumes that the positions of left- and right-eye virtual cameras are unchanged and only the view direction is altered. This assumption is incorrect for actual head motion in VR viewing as the position of both eyes changes with rotation. More advanced positional re-projection techniques rely on screen-space warping, such as Oculus’s asynchronous spacewarp (ASW) [Beeler et al. 2016]. These are known to result in colour bleeding, introduce difficulty in handling translucent and reflective surfaces, and require hole fillings for occluded pixels. Screen-space warping algorithms can either require the application to compute motion vectors for each frame, or to estimate this with an optical flow algorithm. ASW takes the latter approach to promote integration with existing applications, but produces additional optical flow artefacts around repeating patterns and dynamic lighting conditions. Another limitation of re-projection techniques is that there is no bandwidth reduction when transmitting pixels from the GPU to a VR display.

Re-projection techniques are considered a last-resort option in VR rendering, used only to avoid skipped or repeated frames. When the rendering budget cannot be met, lowering the frame resolution is preferred over re-projection [Vlachos 2015].

Other techniques include partial re-projection. Static far-away objects are only affected by rotational movement; such objects can be pre-rendered, and an impostor object can be re-projected in real time [Schauffer 2002]. Such techniques produce promising results, but are heavily scene dependent, might have a significant memory overhead and are not always trivial to integrate with existing pipelines.

### 3.3.4 Non-uniform (foveated) rendering

The non-uniform sensitivity of the eye for different eccentricities (distance from the fovea) has been discussed in Section 2.4. Geisler and Perry [1998] were the first to exploit these limitations by encoding and decoding videos in a foveated multi-resolution pyramid, assigning higher pixel density to lower eccentricities. Guenter et al. [2012] extended this technique to computer graphics content, rendering three layers at different pixel densities, then combining these with a soft stepping function.

In the periphery, orientation resolution is lower than target detection. I.e., the visual system can often detect the presence (or lack) of image information at a given frequency at a given location, but it is not so apt to resolve its orientation. Incorrect orientation is hard to detect. Patney et al. [2016] utilised this insight, and argued that the information lost during down-sampling in foveated rendering can be re-introduced with a post-processing unsharp mask. Such sharpening is likely to introduce detail with the wrong orientation. The authors demonstrated this technique on a VR headset saving up to of the 70% of shading computations.

More recently, deep learning has been applied to reconstruct sparsely-sampled images [Kaplanyan et al. 2018]. The samples follow a blue noise pattern with density corresponding to the sampling density of the visual system. However, the performance impact of their neural network outweighed the potential performance savings.

Foveated rendering algorithms can introduce two types of artefacts: temporal aliasing (flicker), or over-smoothing in the periphery (radial blur, tunnel vision). As the perceptual trade-off between these two is yet unknown, designing a robust and efficient foveation algorithm is challenging. This, and the cost and imprecision of eye trackers have so far prevented wide commercialisation of foveated rendering.

## 3.4 Image and video metrics

Image and video metrics play a crucial role in calibrating, evaluating, and driving novel display techniques. Metrics can be broadly categorised as quality metrics and visibility metrics. In full-reference quality metrics such as PSNR, the model takes a pair of images as input: a reference  $R$  and a test  $T$  image, outputting a single quality value for the entire image. Visibility metrics, on the other hand, produce a distortion map, providing spatial information on the probability of detecting artefacts. Some metrics, such as SSIM produce a distortion map, but it is only the mean value over the image that is shown to correlate with subjective scores [Wang et al. 2003, 2004]. More accurate difference predictions are achieved by white-box models based on knowledge of the human visual system, such as the visual difference predictor (VDP) [Daly 2005], and further improved versions HDR-VDP and HDR-VDP 2 [Mantiuk et al. 2005, 2011]. VDPs are fundamentally multi-scale

models, decomposing the difference between  $R$  and  $T$  into a number of spatial frequency bands as inspired by the multi-resolution model of the CSF discussed in Section 2.4. For full detail of how edge orientation, masking and neural noise is taken into account, please refer to [Mantiuk et al. 2011].

More recent attempts have used convolutional neural networks as a black-box model of vision to improve image difference predictions [Wolski et al. 2018; Ye et al. 2019], pre-trained on the predictions of HDR-VDP 2.

However, these metrics do not take temporal change and chromaticity into account. In Chapter 6, I demonstrate how the multi-scale architecture of VDP can be used to build visual models for detecting colour banding and later in Chapters 7 and 8 for quantifying motion quality.

## 3.5 Summary

The steady improvement of GPUs and display technology allows us to render beyond the spatial frequency limitations of the eye, but there is a long way to go in other dimensions. Insights and models of the visual system can help us understand where exactly the biological and perceptual limits are, and to develop trade-offs between these dimensions. It is often the temporal resolution that is sacrificed to (1) increase bit-depth (dithering), (2) add colour in DLP projectors, (3) enhance resolution or (4) other detail shading quality. Re-projection is a popular method to make up for the reduced motion quality, but is known to fail with translucent and reflective surfaces.

Existing visual metrics have been shown to perform well for images in free-viewing conditions, but there is a lack of metrics for peripheral rendering, colour stimuli, and temporally changing signals. With renewed interest in VR, where motion artefacts are known to be particularly objectionable, it is crucial to focus on the temporal domain more.

In this work, I focus on temporal multiplexing, specifically, on the trade-off between the spatial and temporal domain. In Chapters 5 and 8, I introduce two temporal multiplexing algorithms. Temporal resolution multiplexing (Chapter 5) relies on simple insights into the visual system to reduce the computational cost of rendering while maintaining high perceived quality. In Chapter 8 I propose a visual model for capturing the trade-off between spatial and temporal resolution, and propose an algorithm which adaptively sets the resolution and refresh rate during real-time rendering.

---

---

# CHAPTER 4

---

## DISPLAY PROFILING

*“I think that it is a relatively good approximation to truth – which is much too complicated to allow anything but approximations...”*

*John von Neumann*

*The Mathematician Part 2*

Implementation, testing and psychophysical validation of any perceptual rendering technique rely on a detailed understanding of present display technologies. In this chapter, I briefly describe the operation of OLED and LCD displays, illustrating the different characteristics with physical measurements.

I also fit the popular gain-offset-gamma [Berns 1996] display models for a number of displays, and propose an extension of this for high-refresh-rate LCD monitors. The proposed algorithms in the rest of the dissertation all operate in linear colour spaces, and rely on these models to transform to and from native pixel values.

### 4.1 Displays

Measurements were taken for the following six displays. In the rest of the dissertation I will refer to these displays with the name in the brackets.

- (1) Dell Inspiron 17R 7720 3D laptop display panel,  $1920 \times 1080$ , 15.6”, 120 Hz (Dell)
- (2) Samsung SyncMaster 2233,  $1920 \times 1080$ , 22”, 120 Hz (Samsung)
- (3) Asus ROG SWIFT PG279Q Gaming Monitor,  $2560 \times 1440$ , 27”, G-Sync capable 165 Hz (Asus)

- (4) Huawei Mate 9 Pro in normal and VR mode,  $2560 \times 1440$ , 5.5", 60 Hz, (Huawei, HuaweiVR)
- (5) HTC Vive head-mounted display,  $1080 \times 1200$  per eye,  $110^\circ$  field of view, 90 Hz (Vive)
- (6) Oculus Rift CV1 head-mounted display,  $1080 \times 1200$  per eye,  $95^\circ$  field of view, 90 Hz (Oculus)

## 4.2 Measurement equipments, methods

To gain a better understanding of the spatio-temporal behaviour of state-of-the-art displays, I performed two types of measurements: (1) photometric measurements to map relative pixel luma values to luminance and chromaticity; and (2) high-frequency temporal measurements to investigate the temporal behaviour of displays.

### 4.2.1 Photometric measurements

Luminance and chroma response of each display for stationary stimuli were measured using a Specbos 1211 Spectro-Radiometer with a temporal integration time of 1 second. During measurement, the screen was filled with a uniform colour patch to cancel spatial dithering techniques. The colours ranged from minimal (0) to maximal (1) brightness for all primary components (**R**ed, **G**reen, **B**lue), all secondary colours (**C**yan, **M**agenta, **Y**ellow) and **W**hite.

The spectrometer was placed in a typical viewpoint. For standard desktop displays this was set 50cm away from the panel along the surface normal. To validate alignment with the surface normal, a flat mirror was placed on the screen surface, and the deviation in Specbos's targeting laser's source and reflection was measured. For all measurements this was less than 1mm, which yields a negligible maximum error of  $\pm 0.57^\circ$ . For VR I used an estimated eye position within the eye box and validated that rotating the headset  $\pm 10^\circ$  did not yield significantly different readings (see Figure 4.1).

Measurements were taken in darkened rooms or during the night when ambient luminance fell below the sensor's detection threshold of  $0.1 \text{ cd/m}^2$  [JETI 2010]. To improve the device's reported accuracy (1% luminance reproducibility), each measurement was repeated three times, taking the average value as the final estimate. The order of the measurements was randomised to reduce systematic errors due to heat.

For a complete visualisation of the measurements, please refer to Appendix A. Table 4.1 contains a summary of major display characteristics.





Figure 4.1: Luminance and chrominance measurement setup for HTC Vive. The headset is mounted on a rotating disc to allow measurements from different angles. Specbos is aligned such that the axis of the measurement cone always goes through the centre of the hypothetical eye.

## 4.2.2 Temporal measurements

Specbos correctly establishes the overall perceived colour by using a finite integration time, however, to observe faster temporal variations such as low-persistence and back-light modulation, we need a sensor with higher temporal resolution. With the help of a lab technician, I combined an Arduino Mega board with a photodiode and a negative feedback amplifier using an LM358N op-amp and two  $1\text{M}\Omega$  resistors. Figure 4.2 shows the completed design which accomplished a sampling rate of  $\approx 9\text{kHz}$ . During measurement the photoreceptor was placed within a few centimetres to the display panel. The resultant readings correlate with emitted photons, however, as the change in the signal is more crucial than absolute values, no attempt was made to convert to physical luminance ( $\text{cd}/\text{m}^2$ ). As the temporal profile is not expected to depend on chroma, only a white uniform colour patch was used.

## 4.3 LCD displays

LCD (liquid-crystal displays) is the most popular technology in current flat-panel displays such as desktop and laptop monitors, and smart phones. The LCD layer itself does not emit photons, it merely filters the backlight utilising a pair of polarisers. Applying voltage across the crystals changes the orientation of the molecules, controlling whether each rgb sub-pixel blocks or allows photons to pass through. The filtering layer is imperfect, it

	peak luminance ( $\text{cd}/\text{m}^2$ )	black level ( $\text{cd}/\text{m}^2$ )	dynamic range	technology	backlight
Dell	324.9	0.3	1083:1	LCD	LED
Samsung	192.8	0.2	964:1	LCD	CCFL
Asus	155.2	0.1	1552:1	LCD	LED
Vive	183.32	$0.02^*$	—	OLED	—
Oculus	84.26	$0.01^*$	—	OLED	—
Huawei	371.4	$0.01^*$	—	OLED	—
HuaweiVR	40.5	$0.01^*$	—	OLED	—

Table 4.1: Summary of display characteristics from the luminance measurement. Peak luminance is measured for **White** stimuli; dynamic range is the ratio of maximal over minimal luminance (white level divided by black level). \*: black levels below the reported measurement range. Dynamic range cannot be established in this case. For perfect OLEDs this is expected, as such displays do not emit any photons when the screen is black.

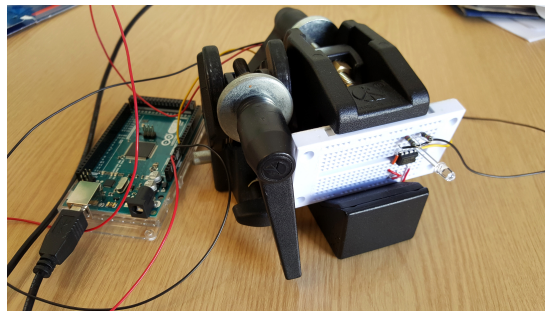
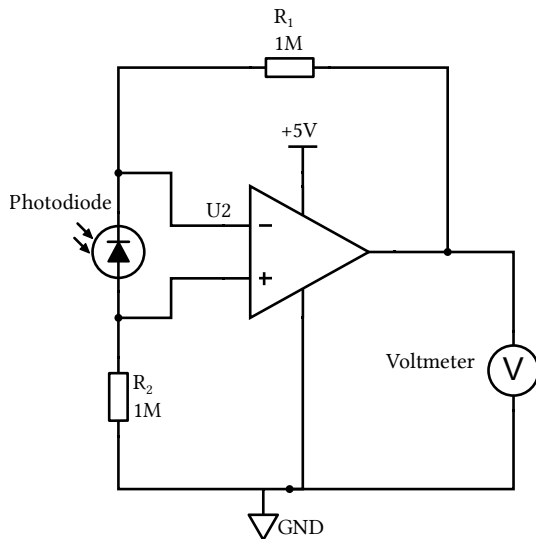


Figure 4.2: High-resolution irradiance measurement sensor. The photocurrent is converted to amplified voltage, then recorded by the Arduino Mega board at  $\approx 10\text{kHz}$ .

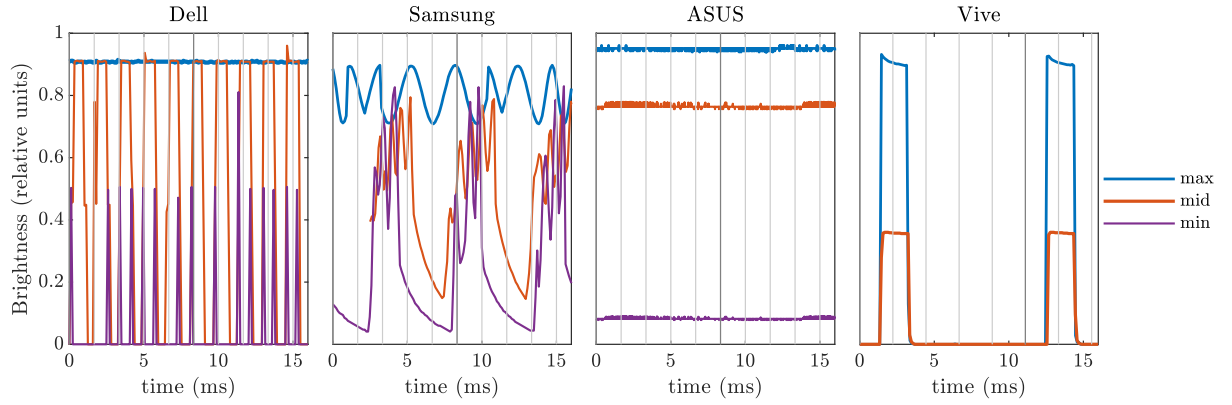


Figure 4.3: Backlight measurements showing different techniques for achieving maximum (max; blue), minimum (min; purple) and medium (mid; red) brightness settings on a monitor. Dell: pulse-width-modulation on LCD-LED display; Samsung: typical CCFL profile with 180 Hz backlight frequency; ASUS: backlight change is unobservable at 10kHz; Vive: daylight vs. night mode alters only peak value, duty cycle remains unchanged on OLED-based VR headset.

cannot fully block the backlight. The amount of light emitted when displaying a black screen is often referred to as the *black level*. The ratio of the brightest white (backlight filtered through the *transparent* LCD layer) and the darkest black (backlight blocked by the LCD layer) is known as the dynamic range or contrast.

### 4.3.1 Backlight

While the light source has been traditionally a cold-cathode fluorescent lamp (CCFL) which provides approximately uniform illumination at all parts of the screen, recent advances in manufacturing technology has introduced LEDs to locally control the backlight. As shown in Table 4.1, LED monitors have a higher dynamic range and often surprisingly high peak luminance values.

Monitors often support a form of brightness adjustments to accommodate viewer preferences in different viewing conditions. Reducing LCD voltages is impractical and often insufficient to adjust brightness levels without sacrificing some of the dynamic range, hence such adjustments usually take place in the backlight. In traditional LCD-CCFL displays this is achieved by dimming the lamp. However, the CCFL light source flickers at relatively low frequencies; in case of the Samsung SyncMaster monitor, a clear 180 Hz modulation signal emerges (Figure 4.3-Samsung). As this particular monitor is updating at 120 Hz, which means that consecutive frames will not receive the same amount of backlight, making such setups unsuitable for delicate temporal multiplexing schemes. LED backlights achieve different brightness levels by flickering the LEDs. For instance, when the display is dimmed in the Dell monitor, the duty cycle of the LEDs are reduced (Figure 4.3-Dell). The emerging square-wave signal has a fundamental frequency  $\approx 600$  Hz,

which is well beyond the visible range of flicker. Backlight modulation in the ASUS monitor was not observable with the 10kHz sensor, making it ideally suited to develop temporal multiplexing techniques.

Low-persistence behaviour can be achieved in LCD monitors by activating the back-light LED for a single burst, for a fraction of the frame time. None of the displays discussed in this chapter were observed to perform such behaviour.

### 4.3.2 Display model

The correspondence between gamma-compressed pixel values, as represented by the computer, and displayed light is usually captured by a display model. The gamma-compressed values are also referred to as *luma* for greyscale images. Without a loss of generality I will assume that on the computer colour is stored as *rgb*, and pixel values are normalised to lie in the 0..1 range. In this dissertation I represent physically displayed light in the CIE-1931 XYZ colour-space unless otherwise stated.

When a display model is invertible, we can define the inverse display model which outputs raw pixel values as a function of desired XYZ values. Note that when feeding XYZ values outside the colour gamut of the display, the inverse display will output values outside the 0..1 range.

The most commonly used invertible model, initially suggested for CRT monitors, is the gain-offset-gamma model [Berns 1996]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r^{\gamma_r} \\ g^{\gamma_g} \\ b^{\gamma_b} \end{bmatrix} \times M_{RGB \rightarrow XYZ} + B, \quad (4.1)$$

where  $XYZ$  is the output light in CIE-XYZ,  $rgb$  is the input pixel luma for red green and blue respectively,  $M_{RGB \rightarrow XYZ}$  is a  $3 \times 3$  transformation matrix,  $\times$  is matrix multiplication, and  $B$  is the three-component black level vector (as measured in CIE-XYZ). The inverse model is then:

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \left( \left( \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - B \right) \times M_{XYZ \rightarrow RGB} \right)^{[1/\gamma_r, 1/\gamma_g, 1/\gamma_b]}, \quad (4.2)$$

where  $M_{XYZ \rightarrow RGB}$  is  $M_{RGB \rightarrow XYZ}^{-1}$ , and power in the square brackets denote a component-wise power operation.

The black level can be measured directly by displaying a black patch on the monitor. Other parameters can be computed by measuring corresponding  $rgb$  and  $XYZ$  values, then solving for the least-squares fit in pixel (luma) space. Display model parameters for

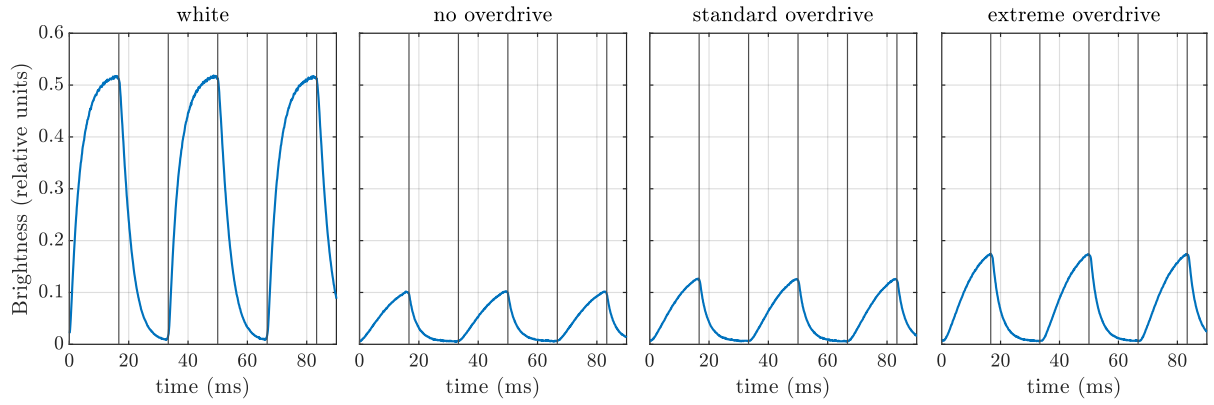


Figure 4.4: LCD crystals switching from black to white (leftmost) and grey (other 3). As crystal switching is not immediate, crystals cannot reach the desired value by the end of the frame time. Overdrive can partially alleviate this problem, switching to a higher luminance value to cancel the effects of switching time. Dark grid lines indicate frame boundaries. Measurements were collected on the ASUS monitor at 60 Hz.

all discussed monitors were measured. Detailed results are listed in Appendix A.

### 4.3.3 Overdrive

The liquid crystals in the recent generation of LCD panels have relatively short response times and offer between 160 and 240 frames a second. However, liquid crystals still require time to switch from one state to another, and the desired target state is often not reached within the time allocated for a single frame. This problem is partially alleviated by over-driving (applying higher voltage), so that pixels achieve the desired state faster (see Figure 4.4).

Switching from one grey-level to another is slower than switching from black-to-white or white-to-black. This non-linear behaviour adds significant complexity to modelling display response, which I address in Section 4.5.

## 4.4 OLED displays

OLED displays do not require a backlight, each pixel emits light depending on the applied voltage. Output luminance can be finely controlled, hence dimming is simply achieved by altering peak luminance as demonstrated in Figure 4.3. When no voltage is applied, each pixel can go completely black, resulting in an extremely high dynamic range. Peak luminance values are also matching or exceeding current LCD generations.

Output transitions are extremely fast compared to traditional LCDs, allowing for a low-persistence mode of operation (Figure 4.5). Such mode is also sometimes labelled VR mode, due to the low-persistence behaviour reducing motion blur and hence simulation

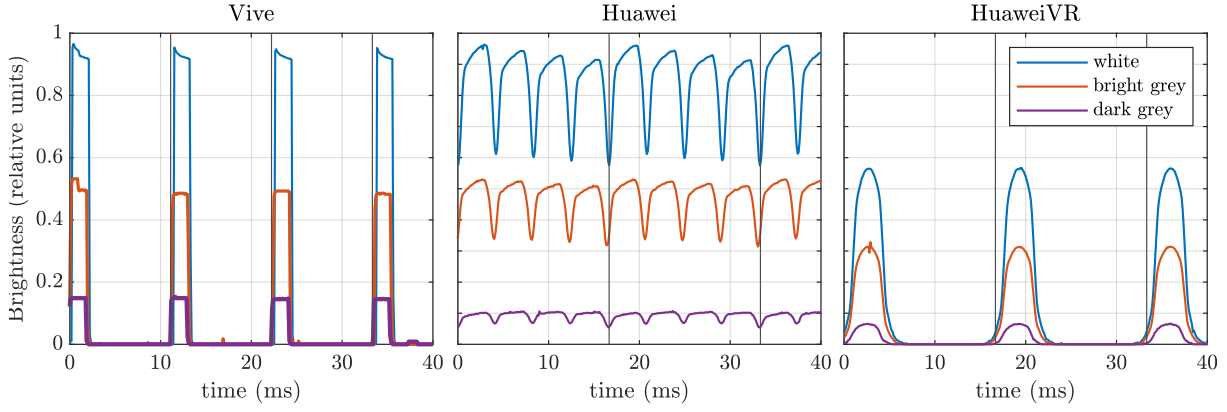


Figure 4.5: Temporal profile of OLED displays. Colours denote different pixel luma values. The low-persistence behaviour can be clearly observed on the HTC Vive with a refresh rate of 90 Hz and a short duty cycle that only lasts  $\approx 18\%$  of the entire frame regardless of pixel luma. Huawei Mate Pro 9 supports two modes of operation: normal mode flickers on 240Hz, repeating each frame four times, whereas VR mode only uses the first window with reduced maximum luminance to avoid perceivable flicker (left). Different luma values are presented by altering the peak value, the shape of the window is unchanged.

sickness. The duty-cycle of low persistence (around 25% for Huawei Mate Pro, less than 18% for the Vive) means that luminance is reduced according to the Talbot-Plateau law. The Huawei Mate Pro 9 flickers on 240Hz, repeating each frame four times in normal mode (overall refresh rate of 60Hz). When a VR application is launched, the display switches to low-persistence mode, where only the first window is used – the screen emits no light in-between. Different luma levels are displayed by changing the height of the peak, leaving the width unaltered. The fast transition speed and high dynamic range makes OLEDs ideal targets for visual science research [Cooper et al. 2013], as well as novel VR and temporal multiplexing schemes. However, manufacturing cost and panel burn-ins have been slowing down wide-spread usage.

## 4.5 High-refresh-rate LCD model

Due to the finite and different rising and falling response times of liquid crystals, we need to consider the previous pixel value when modelling the per-pixel response of an LCD. I used a Specbos 1211 with a 1-second integration time to measure alternating pixel value pairs displayed at different refresh rates on the ASUS monitor. Figure 4.6 illustrates the difference between predicted luminance values (sum of two linear values, estimated by gain-offset-gamma model) and actual measured values. The inaccuracies are quite substantial, especially for low luminance.

To accurately model LCD response, I provide an extension of the gain-gamma-offset display model to account for the pixel value in the previous frame. The forward display

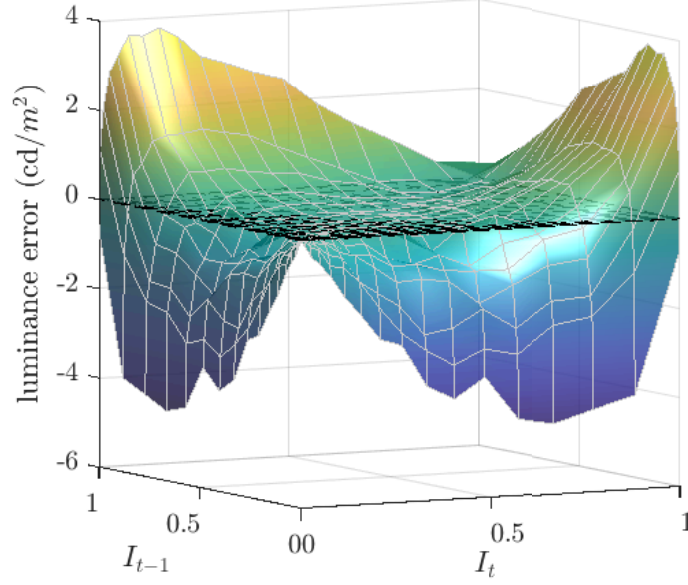


Figure 4.6: Luminance difference between measured luminance value and expected ideal luminance (sum of two consecutive frames) for  $I_t$  and  $I_{t-1}$  pixel values alternating at 165 Hz. My measurements for ASUS ROG Swift P279Q indicate a deviation from the plane when one of the pixels is significantly darker or brighter than the other.

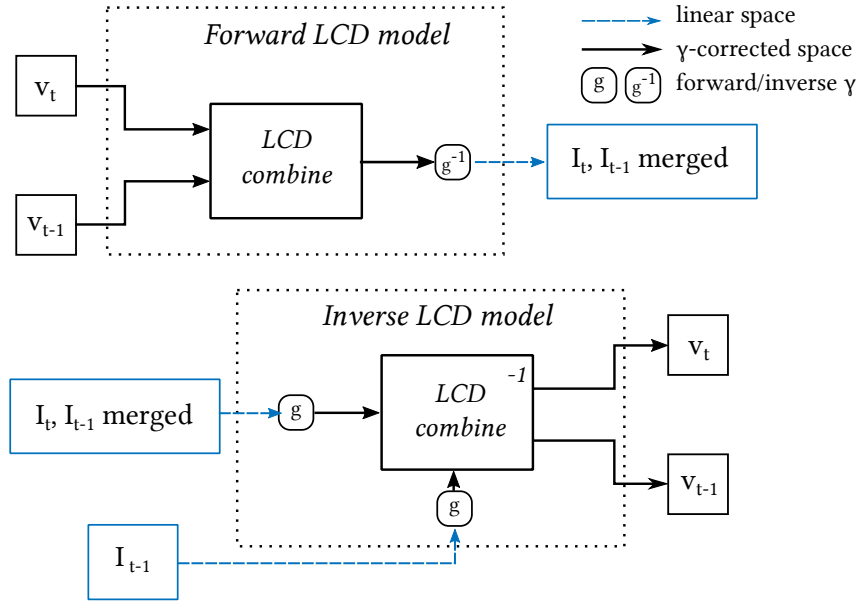


Figure 4.7: Schematic diagram of the extended LCD display model for high-frame-rate monitors. **a)** In the forward model two consecutive pixel values are combined before applying inverse gamma. **b)** The inverse model applies gamma before inverting the LCD combine step. The previous pixel value is provided to find a  $\langle v_t, v_{t-1} \rangle$  pair, where  $v_{t-1}^\gamma \approx I_{t-1}$



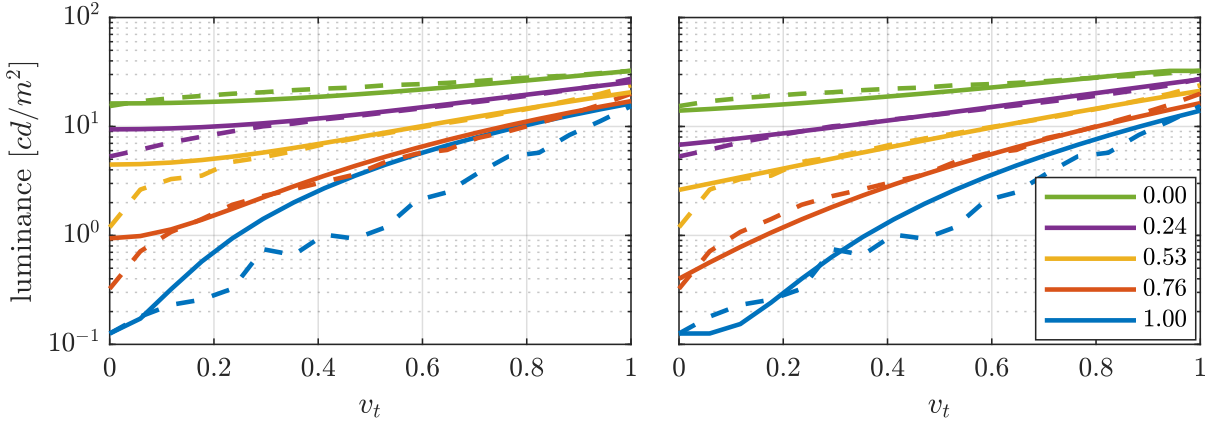


Figure 4.8: Dashed lines: measured display luminance for red primaries ( $v_t$ ), given a range of different  $v_{t-1}$  pixel values (line colours). Solid lines: predicted values without temporal display model (left) and with the proposed temporal model (right).

model, shown in the top of Figure 4.7, contains an additional *LCD combine* block that predicts the equivalent gamma-compressed pixel value, given pixel values of the current and previous frames. Such a relation is well-approximated by a symmetric bivariate quadratic function of the form:

$$M(v_t, v_{t-1}) = p_1(v_t^2 + v_{t-1}^2) + p_2 v_t v_{t-1} + p_3(v_t + v_{t-1}) + p_4, \quad (4.3)$$

where  $M(v_t, v_{t-1})$  is the merged pixel value,  $v_t$  and  $v_{t-1}$  are the current and previous gamma-compressed pixel values and  $p_{1..4}$  are the model parameters. To find the inverse display model, the inverse of the merge function needs to be found. The merge function is not strictly invertible as multiple combinations of pixel values can produce the same merged value. However, since rendering is often in real-time, and we can only control the current and not the previous frame,  $v_{t-1}$  is already given and we only solve for  $v_t$ . If the quadratic equation leads to a non-real solution, or a solution outside the display dynamic range,  $v_t$  needs to be clamped to be within 0..1. To store the residuals from the clamping, we can solve for  $v_{t-1}$ . The difference in prediction accuracy for a single-frame and the proposed temporal display model is shown in Figure 4.8. The parameter values for 165 Hz for each channel (rgb) are as follows:

$$p_{rgb} = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 \\ r : 0.2054 & -0.3433 & 0.4986 & -0.0259 \\ g : 0.2372 & -0.3863 & 0.4932 & -0.0278 \\ b : 0.2601 & -0.4331 & 0.4908 & -0.0253 \end{bmatrix} \quad (4.4)$$

Optimised parameter values for 90 Hz, 120 Hz are quoted in Section A.3.1.



## 4.6 Summary

A thorough understanding of display technologies and reliable models to predict their behaviour is essential to implement and validate novel rendering techniques. In this chapter, I described how photometric measurements were collected and gain-gamma-offset models were fitted to a number of different displays used throughout the rest of the dissertation. I also described how OLED and LCD displays differ, and how the backlight frequency and temporal profile of LCDs can interfere with high-refresh-rate algorithms. To account for some of this, I have proposed a novel (invertible) extension of the gain-gamma-offset model for high-refresh-rate LCD displays.



---

---

## CHAPTER 5

---

# EXPLOITING PERCEPTUAL INSIGHTS: TEMPORAL RESOLUTION MULTIPLEXING

*“And yet a great advantage would be lost, if so simple a law as Weber’s law could not be used as an exact or at least sufficiently approximate basis for psychic measurement; just such an advantage as would be lost if we could not use the Kepler law in astronomy...”*

*Gustav Theodor Fechner  
Elements of psychophysics*

Simple insights into the visual system can prove to be mighty weapons to optimise existing rendering algorithms without introducing visible artefacts. In this chapter I propose an efficient rendering algorithm which reduces both the bandwidth and computational cost exploiting the limited spatio-temporal resolution of the human eye. The following sections are heavily based on the best-journal-paper-award-winning 2019 IEEE Transactions on Visualization and Computer Graphics (TVCG) article titled *Temporal Resolution Multiplexing: Exploiting the limitations of spatio-temporal vision for more efficient VR rendering*.

## 5.1 Introduction

Real-time rendering algorithms are struggling to keep up with the increasingly higher display resolutions and refresh rates. Especially in the context of AR and VR, the sheer number of pixels drawn each second poses a challenge for both computing and transmitting frames from the GPU to the display. The main goal of a robust multiplexing algorithm is to address both of these issues.

As discussed in Chapter 2, the eye has very limited sensitivity to signals changing with both high spatial and temporal frequencies. In other words, fine details flickering at a high refresh rate are not perceivable to human observers, and removing them from the rendering does not result in objectionable artefacts.

In this chapter, I propose a novel technique, Temporal Resolution Multiplexing (TRM), which avoids rendering the high spatio-temporal part of the images by operating on reduced-resolution render targets for every even-numbered frame. This reduces the number of pixels rendered and can be potentially used to reduce the amount of data transferred to the display by 37–49%. TRM then compensates for the contrast loss, making the reduction almost imperceivable. TRM takes advantage of the limitations of the human visual system: the finite integration time that results in fusion of rapid temporal changes, along with the inability to perceive high spatio-temporal frequency signals. An illusion of smooth motion is generated by rendering a low-resolution version of the content for every odd frame, compensating for the loss of information by modifying every even frame. When the even and odd frames are viewed at high refresh rates ( $> 90$  Hz), the visual system fuses them according to the Talbot-Plateau law, and perceives the original full resolution content. The proposed technique, although conceptually simple, requires much attention to details such as overcoming dynamic range limitations, ensuring that potential flicker is invisible, and designing a solution that will save both rendering time and bandwidth. In the following section, I describe these in more detail, then present a psychophysical experiment validating the performance of the technique.

## 5.2 Method

The diagram of the processing pipeline is shown in Figure 5.1. The two high-level blocks in the diagram are *Rendering & encoding*, and *decoding & display*. The two are separated, as they may be realised in different hardware devices: typically rendering is performed on a GPU, and decoding & display is performed by a VR headset. The separation into two parts is designed to reduce the amount of data sent to a display. The optional encoding and decoding steps may involve chroma subsampling, entropy coding or a complete high-efficiency video codec, such as h265 or JPEG XS. All of these bandwidth savings would

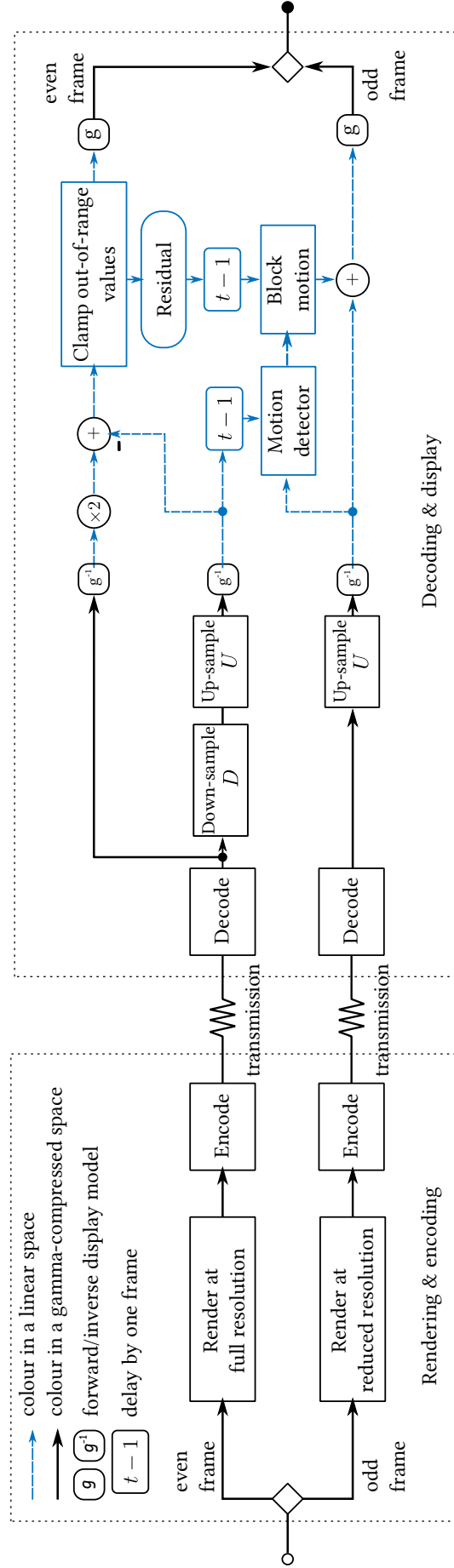


Figure 5.1: The processing diagram for the method. Full- and reduced-resolution frames are rendered sequentially, thus reducing rendering time and bandwidth for reduced resolution frames. Both types of frames are processed so that when they are displayed in rapid succession, they appear the same as the full resolution frames.

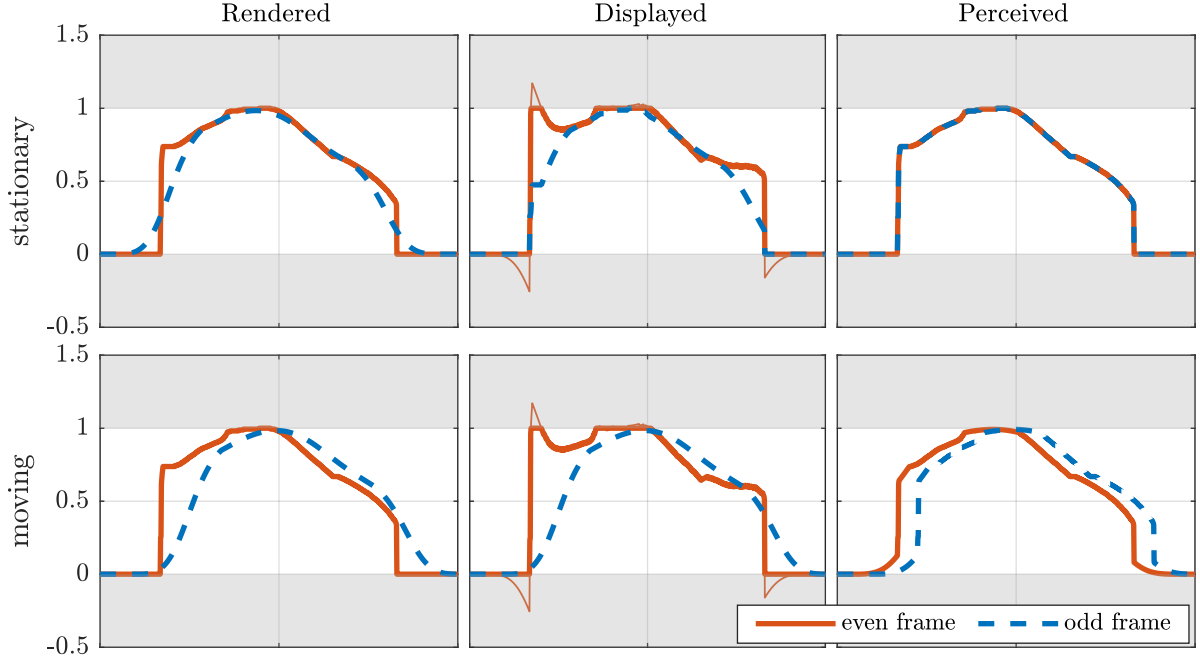


Figure 5.2: Illustration of TRM for stationary (top) and moving (bottom) objects. The two line colours denote odd- and even-numbered frames. After *rendering*, the full-resolution even-numbered frame (continuous orange) needs to be sharpened to maintain high-frequency information. Values lost due to clamping are added to the low-resolution frame (dashed blue), but only whenever the object is not in motion, *i.e. displayed* stationary low-resolution frames are different from the *rendering*, whereas moving ones are identical. Consequently, stationary objects are always perfectly recovered, while moving objects may lose a portion of high-frequency details.

come on top of a 37–49% reduction from TRM.

The top part of Figure 5.1 illustrates the pipeline for even-numbered frames, rendered at full resolution, and the bottom part the pipeline for odd-numbered frames, rendered at reduced resolution. The algorithm transforms those frames to ensure that when seen on a display, they are perceived to be almost identical to the full-resolution and full-frame-rate video. In the next subsections I justify each step of the algorithm (Section 5.2.1), explain how to overcome display dynamic range limitations (Section 5.2.2), and address the problem of phase distortions (Section 5.2.3).

### 5.2.1 Frame integration

The algorithm is designed to be suitable for high refresh rates, specifically, 90Hz or higher, when frame duration is 11.1 ms or less. A pair of such frames lasts approx. 22.2 ms, which is short enough to fit within the range in which the Talbot-Plateau law holds. Consequently, the perceived stimulus is the average of two consecutive frames, one containing mostly low frequencies (reduced resolution) and the other containing all frequencies. Let

us denote the upsampled reduced-resolution (odd) frame at time instance  $t$  with  $\alpha_t$ :

$$\alpha_t(x, y) = (U \circ i_t)(x, y) , \quad t = 1, 3, \dots \quad (5.1)$$

where  $U$  is the up-sampling operator,  $i_t$  is a low-resolution frame and  $\circ$  denotes function composition. Up-sampling in this context means interpolation and increasing sampling rate. Down-sampling in this context is defined as the application of an appropriate low-pass filter and resolution reduction. Note that  $i_t$  must be represented in linear colorimetric values (not gamma compressed); for this I rely on the display models introduced in Chapter 4. The following equations only consider luminance here, but the same analysis applies to linearly-encoded rgb or CIE XYZ colour channels. The initial candidate for the all-frequency even frame, compensating for the lower resolution of the odd-numbered frame, is denoted by  $\beta$ :

$$\beta_t(x, y) = 2I_t(x, y) - (U \circ D \circ I_t)(x, y) , \quad t = 2, 4, \dots \quad (5.2)$$

where  $D$  is a down-sampling function that reduces the size of frame  $I_t$  to that of  $i_t$  ( $i_t = D \circ I_t$ ), and  $U$  is the up-sampling function, the same as that used in Equation 5.1. When an image is static ( $I_t = I_{t+1}$ ), the eye integrates the frames according to the Talbot-Plateau law, with the perceived image being:

$$\alpha_t(x, y) + \beta_{t+1}(x, y) = 2I_t(x, y) . \quad (5.3)$$

Therefore, we perceive the image  $I_t$  at its full resolution and brightness (the equation is the sum of two frames and hence  $2I_t$ ). A naïve approximation of  $\beta_t(x, y) = I_t(x, y)$  would result in a loss of contrast for sharp edges; computing a compensated image  $\beta_t$  is hence a necessary step that prevents the rendered animation from appearing blurry.

The top row in Figure 5.2 illustrates rendered low- and high-frequency components (1st column), compensation for missing high frequencies (2nd column), and the perceived signal (3rd column), which is identical to the original signal if there is no motion. However, what is more interesting and non-obvious is that we will see a correct image even when there is movement in the scene. If there is movement, it is most likely caused by an object or camera motion. In both cases, the gaze follows an object or scene motion (see Section 2.6.1), thus stabilising the image on the retina. As long as the image is fixed, the eye will see the same object at the same retinal position and Equation 5.3 will be valid. Therefore, as long as the change is due to rigid motion trackable by SPEM, the perceived image corresponds to the high-resolution frame  $I$ .

### 5.2.2 Overshoots and undershoots

The decomposition into low- and high-resolution frames  $\alpha$  and  $\beta$  is not always straightforward as the high resolution frame  $\beta$  may contain values that exceed the dynamic range of a display. As an example, let us consider the signal shown in Figure 5.2 and assume that our display can reproduce values between 0 and 1. The compensated high-resolution frame  $\beta$ , shown in orange, contains values that are above 1 and below 0, which I will refer to as overshoots and undershoots. Clamping the “orange” signal to the valid range, the perceived integrated image will lose some high-frequency information and will be effectively blurred. This section explains how this problem can be reduced to the point that the loss of sharpness is imperceptible.

For stationary pixels, overshoots and undershoots do not pose a significant problem. The difference between an enhanced even-numbered frame  $\beta_t$  (Equation 5.2) and the actually displayed frame, altered by clamping to the display dynamic range, can be stored in a *residual buffer*  $\rho_t$ . The values stored in the residual buffer are then added to the next low resolution frame:  $\alpha'_{t+1} = \alpha_{t+1} + \rho_t$ . If there is no movement, adding the residual values restores missing high frequencies and reproduces the original image. However, for pixels containing motion, the same approach would introduce highly objectionable ghosting showing as a faint copy of sharp edges at the previous frame locations. Better animation quality is achieved if the residual is ignored for fast-moving objects. This introduces a small amount of blur for a rare occurrence of high-contrast moving objects, but such blur is almost imperceptible due to the imperfect nature of SPEM (Section 2.6.1) and motion sharpening (see Section 2.6.4). Therefore a weighing mask seems suitable when adding the residual to the odd-numbered frame:

$$\alpha'_{t+1}(x, y) = \alpha_{t+1} + w(x, y) \rho_t(x, y), \quad (5.4)$$

where  $\alpha'(x, y)$  is the final displayed odd-numbered frame. For  $w(x, y)$  let us first compute the Michelson contrast [Kukkonen et al. 1993] between consecutive frames as an indicator of motion:

$$c(x, y) = \frac{|U \circ D \circ I_{t-1}(x, y) - U \circ i_t(x, y)|}{U \circ D \circ I_{t-1}(x, y) + U \circ i_t(x, y)} \quad (5.5)$$

then apply a soft-thresholding function inspired by the Weibull function [1951] :

$$w(x, y) = \exp(-(c_t(x, y))/s_1)^{s_2}), \quad (5.6)$$

where  $s_1$  and  $s_2$  are adjustable parameters controlling the sensitivity to motion – thereby driving the trade-off between ghosting and blurring. In practice I observed that  $s_1 = 0.46$  and  $s_2 = 3$  provided good results for a range of content on a number of displays.

The visibility of blur for moving objects can be further reduced if up-sampling and



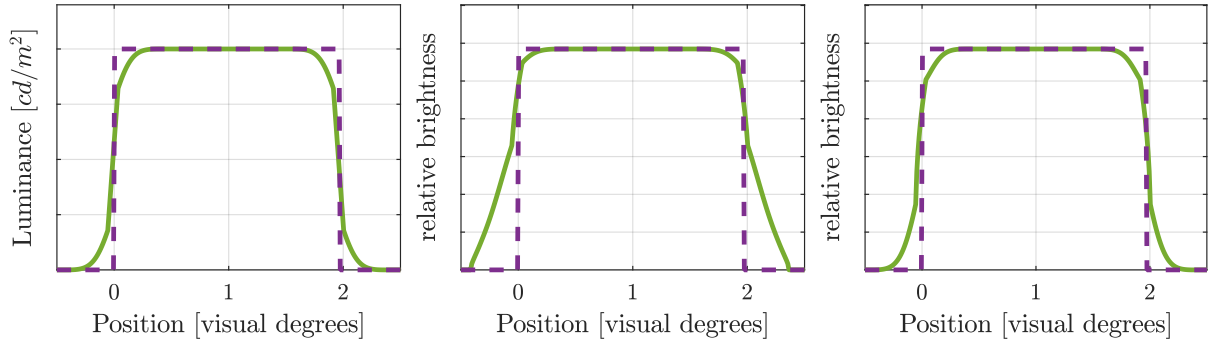


Figure 5.3: The image of a moving square integrated on the retina for the original animation (dashed line) and after applying TRM (solid line). Due to dynamic range limitations, edges can appear blurry. *Left*: In linear luminance space low- and high-luminance artefacts are equally sized; however, such representation is misleading, as brightness perception is non-linear. *Centre*: better estimation of perceived signal using Stevens’s brightness, where artefacts in the dark regions are shown to be more noticeable. *Right*: TRM performs sampling in  $\gamma$ -compressed space, the perceptual impact of artefacts are balanced.

down-sampling occurs in an appropriate colour space. Perception of luminance change is strongly non-linear; blur introduced in dark regions tends to be more visible than in bright regions. The visibility of blur can be more evenly distributed between dark and bright pixels if up-sampling and down-sampling operations are performed in a gamma-compressed space, as shown in Figure 5.3. A cube root-function is considered to be a good predictor of brightness (Stevens’s brightness), and is commonly used in uniform colour spaces, such as CIE Lab and CIE Luv. However, the standard sRGB colour space with gamma  $\approx 2.2$  is sufficiently close to the cube root ( $\gamma = 3$ ) and, since the rendered and transmitted data is likely to be already in that space, it provides a computationally efficient alternative.

### 5.2.3 Phase distortions

A naïve rendering of frames at reduced resolution without anti-aliasing results in a discontinuity of phase changes for moving objects, which reveals itself as juddery motion. A frame that is rendered at lower resolution and up-sampled is not equivalent to the same frame rendered at full resolution and low-pass filtered, as it is not only missing information in high spatial frequencies, but also lacks accurate phase information.

In practice, the correct phase can be reintroduced by utilising hardware-accelerated anti-aliasing, such as multi-sampled anti-aliasing (MSAA) [Carpenter 1984]. Further improvements in quality can be achieved with custom resolve filters (Gaussian or Lanczos) if supported by hardware [Pettineo 2015]. Alternatively, the low-resolution frame can be low-pass filtered to achieve similar results.

In my experiments I used a Gaussian filter with  $\sigma = 2.5$  pixels for both the down-

sampling operator  $D$  and for MSAA resolve. Up-sampling was performed as bilinear interpolation, as it is fast and supported by GPU texture samplers.

#### 5.2.4 Resolution reduction vs refresh rate (Experiment 5.1)

To estimate how much computation can be saved with TRM, I first establish the amount of down-sampling which can be applied without introducing any visible artefacts. I define the resolution reduction factor as the ratio of the rendered resolution and the original display resolution. I measure the maximal imperceivable reduction factor as a function of display refresh rate. More specifically, imperceivable means indistinguishability from standard rendering.

##### Setup:

The animation sequences were shown on the ASUS monitor ( $2560 \times 1440$ , 27"; for full details see Chapter 4). This display, unlike any OLED displays found in VR headsets, allows for fine control over refresh rate. The viewing distance was fixed at 75 cm using a headrest, resulting in the angular resolution of 56 ppd. Custom OpenGL software was used to render the sequences in real-time, with or without TRM. The monitor was driven by a PC equipped with an Intel i7-7700 processor and NVIDIA GeForce GTX 1080 Ti GPU.

##### Stimuli:

In each trial participants saw two short animation sequences one after another, one of them rendered using TRM, the other rendered at the full resolution. Each animation lasted 5-10 s with an average duration of 6 s. Both sequences were shown at the same refresh rate. Figure 5.4 shows a thumbnail of the four animations used in the experiment. The animations contained moving *Chequered Circles*, scrolling *Text*, panning of a *Panorama* and a 3D model of a *Sports hall*. The two first clips were designed to provide an easy-to-follow object with high contrast; the two remaining clips tested the algorithm on rendered and camera-captured scenes. *Sports hall* tested more of a game-like setup by introducing user interaction, letting users rotate the camera with a mouse. The other sequences were pre-recorded. In the *Panorama* scene, the image panned with constant speed.

The animations were displayed at four refresh rates:  $\{100, 120, 144, 165\}$  Hz. Lower refresh rates could not be tested because the display did not natively support 90 Hz, and flicker was visible at lower refresh rates. In this experiment, G-Sync was always disabled to prevent any temporal aliasing artefact from the G-Sync control circuit.



Figure 5.4: Stimuli used for Experiment 5.1.

### Task:

The goal of the experiment was to find the threshold reduction factor at which the observers could notice the difference between TRM and standard rendering with 75% probability. An adaptive QUEST procedure, as implemented in Psychophysics Toolbox extensions [Brainard 1997], was used to sample the continuous scale of reduction factors and to

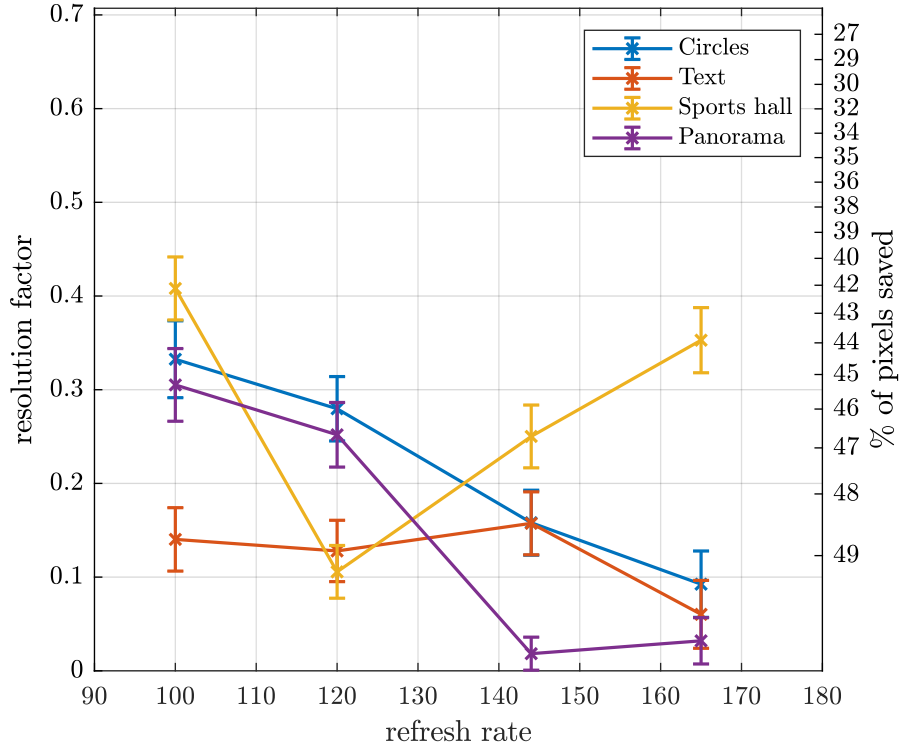


Figure 5.5: Result of Experiment 5.1: finding the smallest resolution reduction factor for four scenes and four display refresh rates. As the reduction is applied to both horizontal and vertical dimensions, the percentage of pixels saved over a pair of frames is computed as  $(1 - r^2)/2 \times 100$ .

fit a psychometric function. The order of trials was randomised so that 16 QUEST procedures were running concurrently to reduce the learning effect. In each trial the participant was asked to select the sequence that presented *better motion quality*. They had an option to re-watch the sequences (in case of lapse of attention), but were discouraged from doing so. Before each session, participants were briefed about their task both verbally and in writing. The briefing explained the motion quality factors (discussed in Section 2.6.3) and was followed by a short training session, in which the difference between 40 Hz and 120 Hz was demonstrated.

### Participants:

Eight paid participants aged 18 – 35 took part in the experiment. All had normal or corrected-to-normal full colour vision.

### Results:

The results in Figure 5.5 show a large variation in the reduction factor from one animation to another. This is expected as motion velocity and contrast varied in this experiment, while both factors strongly affect motion quality. For all animations, with the exception of

the *Sports hall*, the resolution of odd-numbered frames can be further reduced for higher refresh-rate displays. *Sports hall* was an exception in that participants chose almost the same reduction factor for both the 100 Hz and 165 Hz display. Post-experiment interviews revealed that the observers used the self-controlled motion speed and sharp edges present in this rendered scene to observe slight variation in sharpness. Note that this experiment tested discriminability, which results in a conservative threshold for ensuring same quality. That means that such small variations in sharpness, though noticeable, are unlikely to be objectionable in practical applications.

Overall, the experiment showed that a resolution factor of 0.4 or more produces animation that is indistinguishable from rendering frames at the full-resolution. Stronger reduction could be possible for high-refresh displays, however, a 0.4 resolution factor already corresponds to saving 42% of the pixel computations. Further savings towards 50% although possible, are probably not worth risking the introduction of artefacts.

### 5.3 Comparison with other techniques

In this section I discuss other methods intended for improving motion quality or reducing image transmission bandwidth and compare them to the proposed algorithm. Table 5.1 provides a list of common techniques that could be used to save on rendering cost.

Table 5.1: Comparison of alternative techniques.

	Peak luminance	Motion Blur	Flicker	Artefacts	performance saving
Re-projection	100%	reduced	none	re-projection artefacts	varies; 50% max.
Half frame rate	100%	strong	none	judder	50%
Interlace	50%	reduced	moderate	combing	50%
BFI	50%	reduced	severe	none	50%
NCSFI	100%	reduced	mild	ghosting	50%
TRM (proposed)	100%	reduced	mild	minor	37–49%

The simplest way to halve the transmission bandwidth is to *halve the frame rate*. This obviously results in non-smooth motion and severe hold-type blur. *Interlacing* (odd and even rows are transmitted in consecutive frames) provides a better way to reduce bandwidth. Setting missing rows to black can reduce motion blur. Unfortunately, this will reduce peak luminance by 50% and may result in visible flicker, aliasing and combing artefacts. Hold-type blur can be reduced by inserting a black frame every other frame (*black frame insertion* — *BFI*), or backlight flashing [Feng 2006]. This technique, however, is prone to causing severe flicker and also reduces peak display luminance (see Section 7.3

for an illustration of this). Nonlinearity compensated smooth frame insertion (*NCSFI*) [Chen et al. 2006] relies on a similar principle as TRM, and displays sharpened and blurred frames. The difference is that every pair of blurred and sharpened frames is generated from a single frame (from 60 Hz content). The method saves 50% on computation and does not suffer from reduced peak brightness, but results in ghosting at higher speeds, as demonstrated in Experiments 5.2 and 5.3.

Didyk et al. [2010b] demonstrated that up to two frames could be morphed from a previously rendered frame. They approximate scene deformation with a coarse grid that is snapped to the geometry and then deformed in consecutive frames to follow motion trajectories. Morphing can obviously result in artefacts, which the authors avoid by blurring morphed frames and then sharpening fully rendered frames. In that respect, the method takes advantage of similar perceptual limitations as TRM and NCSFI. Re-projection methods (Didyk et al. [2010b], ASW [Beeler et al. 2016]), however, are much more complex than TRM and require a motion field, which could be expensive to compute, reducing the performance saving. Such methods have limitations handling transparent objects, specularities, dis-occlusions, changing illumination, motion discontinuities and complex motion parallax. I argue that rendering a frame at a reduced resolution (as done in TRM) is both a simpler and more robust alternative. Although minor loss of contrast could occur around high-contrast edges such as in Figure 5.3; in Experiment 5.2 and Experiment 5.3 I demonstrate that the failures of a state-of-the-art re-projection technique, ASW, produce much less preferred results than TRM. Moreover, re-projection cannot be used for efficient transmission as it would require transmitting motion fields, thus eliminating potential bandwidth savings.

## Fourier analysis with the window of visibility

To further distinguish the proposed approach from previous methods, I use the window of visibility analysis (Section 2.6.2) considering the spatio-temporal behaviour of a thin, vertical line moving with constant speed from left to right. Such a simplistic animation has proved to be an excellent visualisation tool, and as the discrete approximation of a Dirac delta, poses a good challenge for the compared techniques. Figure 5.6 shows how a single row of such a stimulus changes over time when presented using different techniques. The plot of position vs. time forms a straight line for a real-world motion, which is not limited by frame rate (top row, 1st column). But the same motion forms a series of vertical line segments on a 60 Hz OLED display, as the pixels must remain constant for  $1/60$ -th of a second. When the display frequency is increased to 120 Hz, the segments become shorter. The second column shows the stabilised image on the retina assuming that the eye perfectly tracks the motion. The third column shows the image integrated over time according to the Talbot-Plateau law.

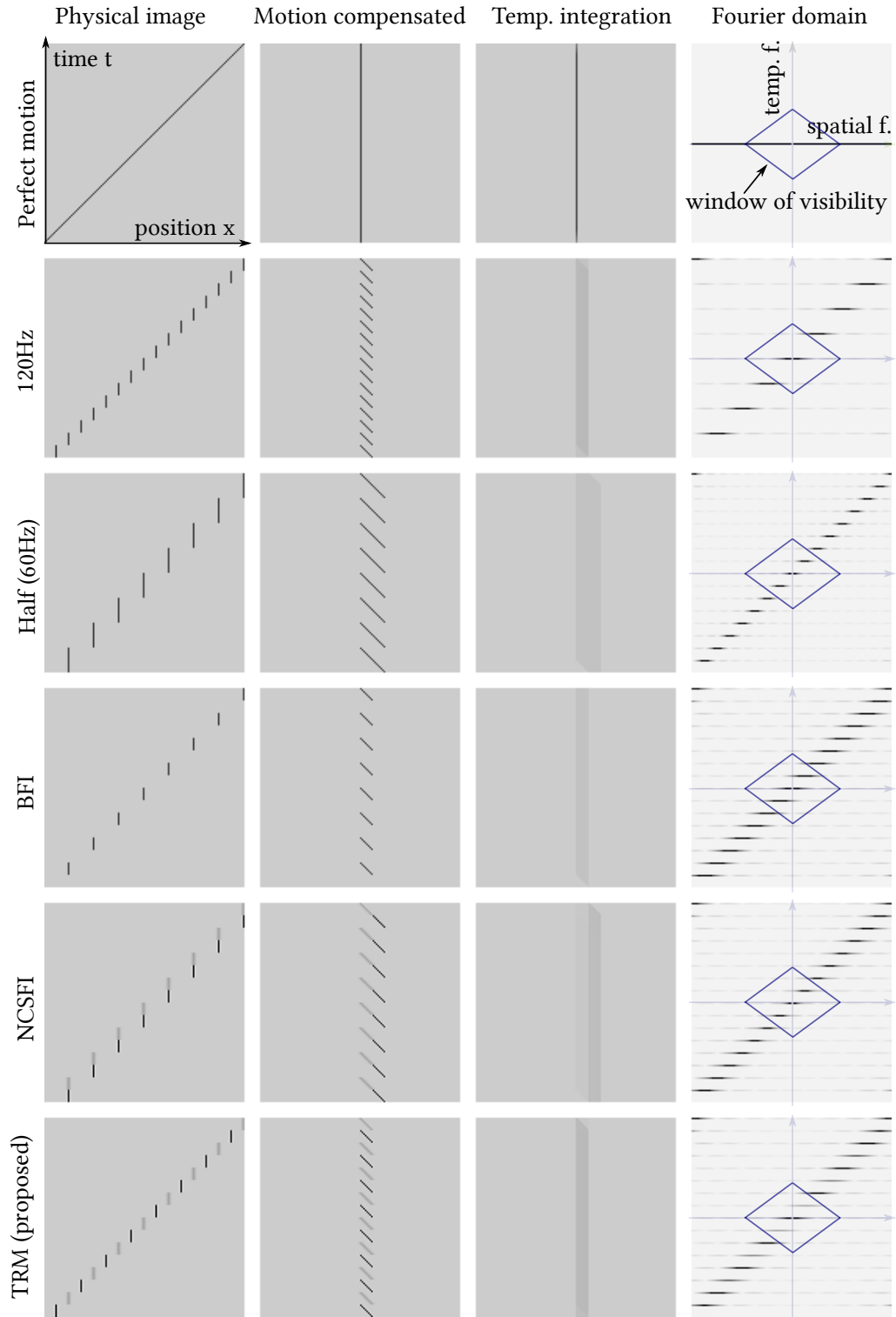


Figure 5.6: A simple animation consisting of a thin, vertical line moving from left to right as seen in real-world (top row), and using different display techniques (remaining rows). The columns illustrate the physical image (1<sup>st</sup>), the stabilised image on the retina (2<sup>nd</sup>) and the image integrated by the visual system (3<sup>rd</sup>). The 4<sup>th</sup> column shows the 2<sup>nd</sup> column in the Fourier domain, where the diamond shape indicates the range of spatial and temporal frequencies visible to the human eye.

60 Hz animation appears more blurry than the 120 Hz animation (see 3rd column) mostly due to a hold-type blur. The three bottom rows compare three techniques aiming to improve motion quality, including ours. The black frame insertion (BFI) reduces the blur to that of 120 Hz without the need to render an image 120 frames per second, but it also reduces the brightness of an image by half. NCSFI [Chen et al. 2006] does not suffer from reduced brightness and also reduces hold-type blur, but to a lesser degree than BFI. TRM (bottom row) has all the benefits of NCSFI but achieves stronger blur reduction, on par with the 120 Hz video.

Further advantages of the proposed technique are revealed by analysing the animation in the frequency domain. The fourth column in Figure 5.6 shows the Fourier transform of the motion-compensated image (2nd column). The blue diamond shape represents the range of visible spatial and temporal frequencies, following the stCSF shape from Figure 2.4-left. The perfectly stable *physical image* of a moving line (top row) corresponds to the presence of all spatial frequencies in the Fourier domain (the Fourier transform of a Dirac peak is a constant value). Motion appears blurry on a 60 Hz display and hence we see a short line along the  $x$ -axis, indicating the loss of higher spatial frequencies. More interestingly, there are a number of aliasing copies of the signal in higher temporal frequencies. Such aliasing copies reveal themselves as non-smooth motion (crawling edges). The animation shown on a 120 Hz display (3rd row) reveals less hold-type blur (longer line on the  $x$ -axis) and it also puts aliasing copies further apart, making them potentially invisible. BFI and NCSFI result in a reduced amount of blur, but temporal aliasing is comparable to a 60 Hz display. TRM reduces the contrast of every second alias, thus making them much less visible. Therefore, although other methods can reduce hold-type blur, only the proposed method can improve the smoothness of motion.

## 5.4 Applications

In this section, I present three different use-cases of TRM. The applications are validated with a set of psychophysical experiments described in Experiments 5.2, 5.3, and 5.4.

### 5.4.1 Virtual reality

To better distribute rendering load over frames in stereo VR, one eye is rendered at full resolution and the other eye at reduced resolution; then, we swap the resolutions of the views in the following frame. Such alternating binocular presentation will not result in higher visibility of motion artefacts than the corresponding monocular presentation. The reason is that the sensitivity associated with disparity estimation is much lower than the sensitivity associated with luminance contrast perception, especially for high spatial and temporal frequencies [Hoffman et al. 2011].



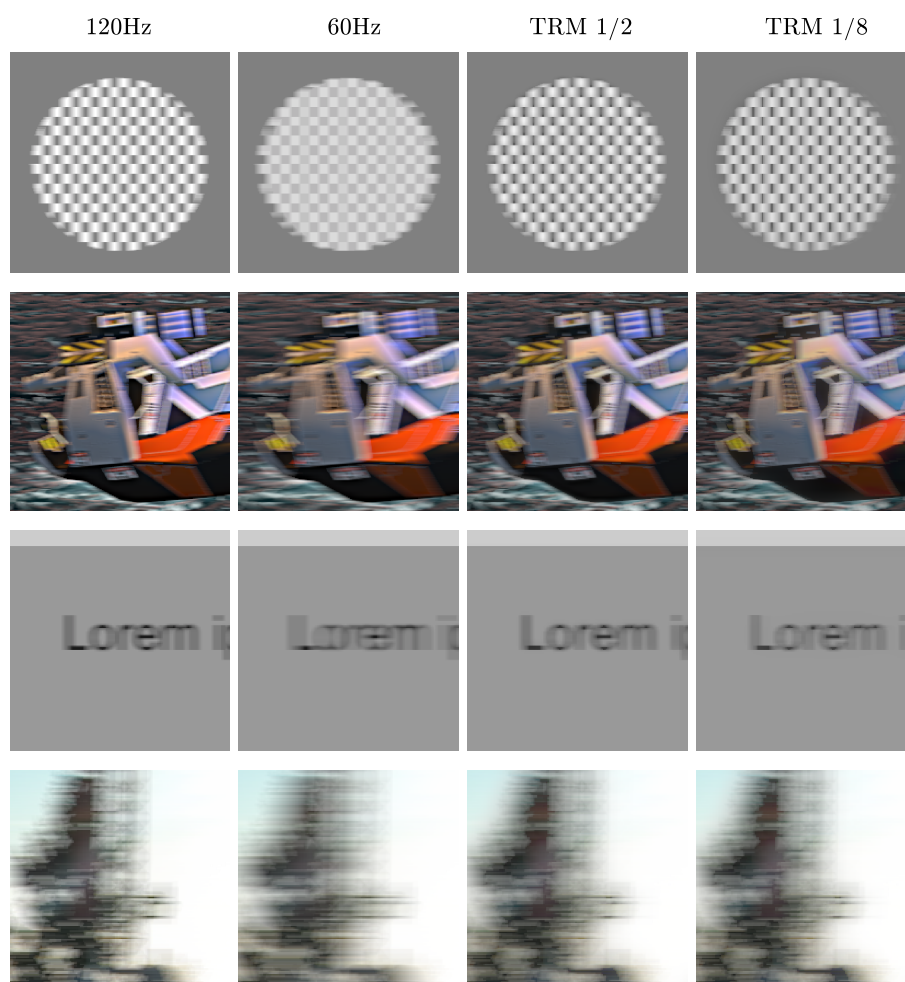


Figure 5.7: Simulation of perceived video frames. At full frame rate (120 Hz) the stimulus looks sharper than for half frame rate (60 Hz). With TRM low-frequency blur is eliminated. The reduction in contrast for high-frequency signal is usually unnoticeable for moving objects.

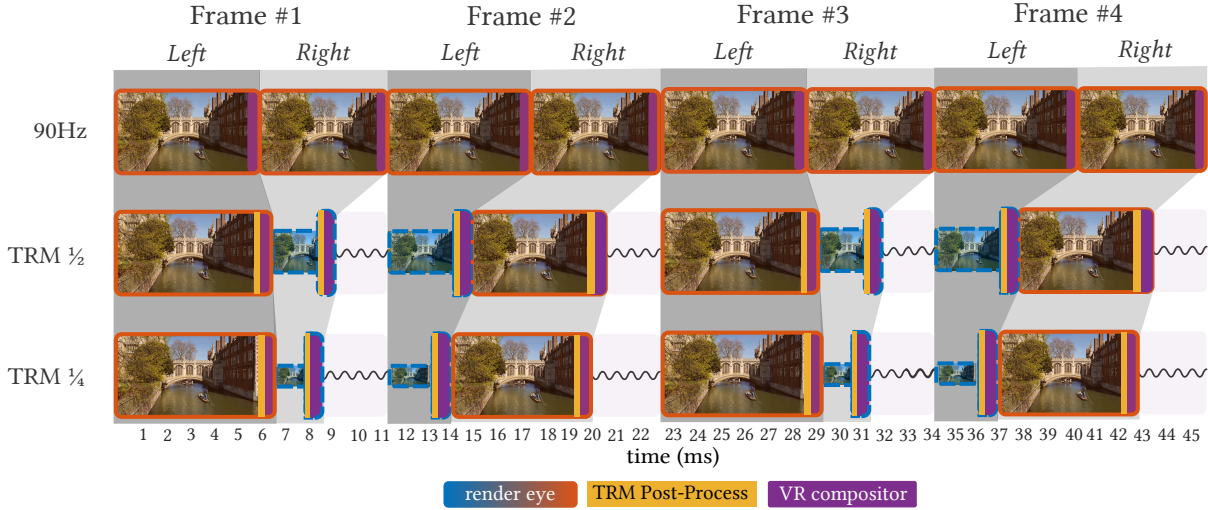


Figure 5.8: Measured performance of 90 Hz full resolution rendering on HTC Vive for four consecutive frames averaged over 1500 samples (top); compared with the proposed TRM method with  $\frac{1}{2}$  and  $\frac{1}{4}$  resolution reduction. Unutilised time periods (shown as wavy lines) could be used to load resources or compute additional visual effects or geometry. Post-processing and VR composition (yellow and purple) take less time than rendering. For each frame, left and right eyes are drawn consecutively. Red borders indicate full-resolution rendering, blue borders indicate reduced-resolution rendering. Image sizes in the illustration are proportionate to the rendered resolution.

Another important consideration is whether the fusion of low- and high-resolution frames happens before or after binocular fusion. Studies on binocular flicker [Perrin 1954] suggest that while most of the flicker fusion is monocular, there is also a measurable binocular component, also known as the Sherrington effect. As a result, off-phase flicker between the two eyes has been shown to be less visible than on-phase flicker. This is actually beneficial for TRM, as long as high- and low-resolution frames are presented to different eyes. Indeed, observers reported that flicker is less visible in a binocular presentation on a VR headset.

Reducing the resolution of one eye can reduce the number of pixels rendered by 37–49%, depending on the resolution reduction. Informal experiments established that a reduction of  $\frac{1}{2}$  (37.5% pixel saving) produces good-quality rendering on the HTC Vive. I measured the performance of TRM in a fill-rate-bound football scene (Figure 5.9 bottom) with procedural texturing, reflections, shadow mapping and dynamic lighting. The light count was adjusted to fully utilise the 11ms frame time on the desktop PC (HTC Vive, Intel i7-7700 processor and NVIDIA GeForce GTX 1080 Ti GPU). As Figure 5.8 indicates, a 19-25% speed-up was observed for an unoptimised OpenGL and OpenVR-based implementation. More optimised applications, especially ones relying on fragment-bound effects such as ray tracing, hybrid rendering [Purcell et al. 2005] and parallax occlusion mapping [Tatarchuk 2006] could benefit even more.

A pure software implementation of TRM can be easily integrated into existing rendering pipelines as a post-processing step. The only significant change in the existing pipeline is the introduction of the stage which alternates between full- and reduced-resolution render targets. In practice, available game engines such as Unreal Engine or Unity either support resizeable render targets or allow light-weight alteration of the viewport through their scripting infrastructure. When available, resizeable render targets are preferred to avoid MSAA resolves in unused regions of the render target, and to prevent bleeding artefacts around the edges of the viewport. Validation results on two VR headsets is presented in Experiments 5.2 and 5.3.

### 5.4.2 High-refresh-rate monitors

The same principle can be applied to high-refresh-rate gaming monitor; utilising savings from resolution reduction to render games at a higher quality or on lower-tier PCs. The technique could also be potentially used to reduce bandwidth for transmission of high-frame-rate video from cameras. However, the difference between 120 Hz and 60 Hz is noticeable mostly for very high angular velocities, such as those experienced in VR and first-person games. The benefit of high refresh rates is more difficult to observe for traditional video content. Validation results for high-refresh-rate desktop monitor is presented in Section 4.

### 5.4.3 Portable devices

With the anticipated dawn of high-refresh-rate embedded devices, TRM could be also implemented on smart phones and tablets to improve text readability while scrolling. The reduced rendering load of TRM is expected to result in lower power consumption and consequently longer battery life when compared to naïve rendering. Latest devices, such as the iPad Pro, are just beginning to add support for >90 Hz refresh rates<sup>1</sup>.

## 5.5 Validation

I conducted three psychophysical experiments to validate the proposed applications

### 5.5.1 Virtual Reality (Experiments 5.2 and 5.3)

The VR validation of the proposed technique is performed in Experiments 5.2 and 5.3, comparing TRM with baseline rendering, and two alternative techniques: NCSFI and state-of-the-art re-projection (ASW).

---

<sup>1</sup>Apple press release — <https://www.apple.com/uk/newsroom/2018/10/new-ipad-pro-with-all-screen-design-is-most-advanced-powerful-ipad-ever>

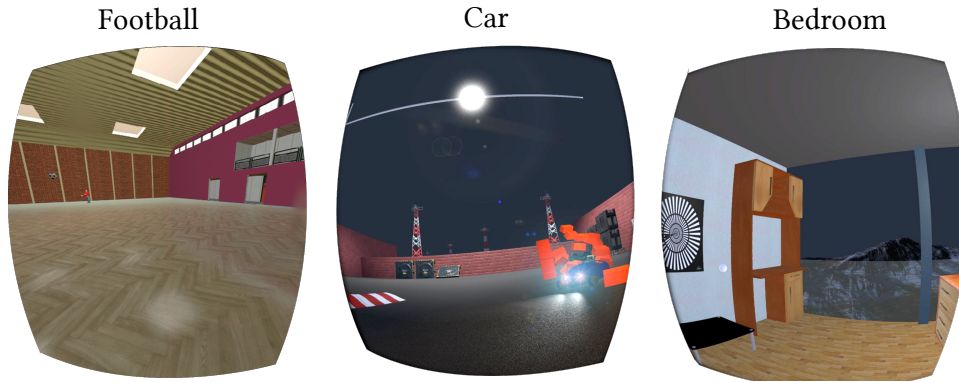


Figure 5.9: Stimuli used for validation in the VR experiments

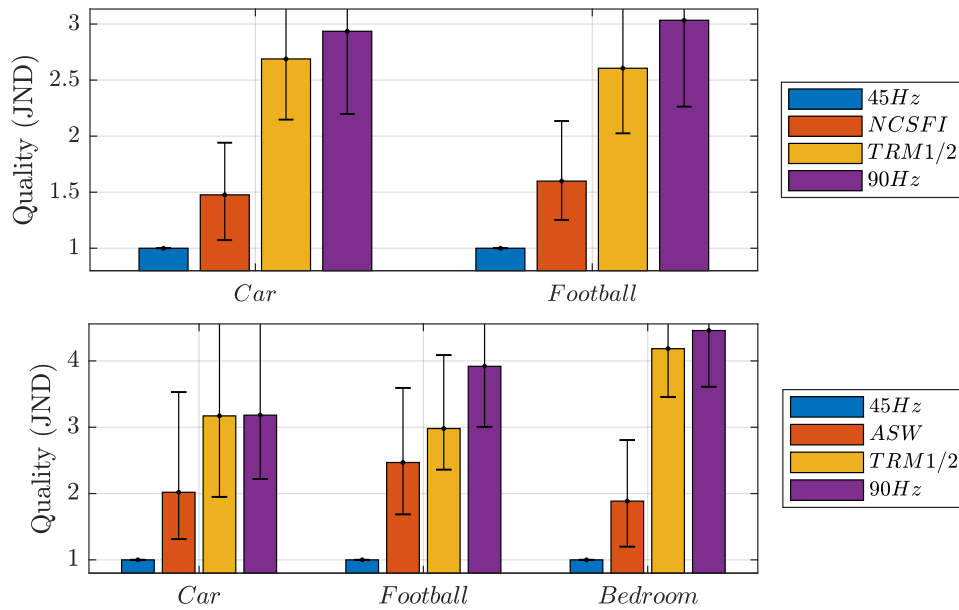


Figure 5.10: Results of Experiment 5.2 on the HTC Vive (top) and Experiment 5.3 on the Oculus Rift (bottom). Error bars denote 95% confidence intervals. The measured quality difference between each pair of techniques is statistically significant, with the exceptions of *TRM* vs. *90 Hz* and *45 Hz* vs. *ASW*.

### Setup:

Two high-end off-the-shelf VR headsets were used running on 90 Hz – HTC Vive and Oculus Rift CV1 for Experiments 5.2 and 5.3 respectively. Participants were asked to perform the experiment on a swivel chair for stability, but were encouraged to move their heads around. State-of-the-art re-projection technique, ASW, is unfortunately not implemented for the HTC Vive, so in Experiment 5.2 I only tested TRM against baseline renderings and NCSFI. In Experiment 5.3, I replaced NCSFI with the latest ASW implementation on the Oculus Rift. I decided to not include NCSFI on the Oculus experiment to avoid over-tiring participants.

### Stimuli:

In each trial the observer was placed in two brief (10s each) computer-generated environments, identical in terms of content, but rendered using one of the following five techniques: (1) 90 Hz full refresh rate, (2) 45 Hz halved refresh rate, duplicating each frame (3) TRM with a  $\frac{1}{2}$  down-sampled render target for every other frame (4) non-linearity compensated smooth frame insertion (NCSFI) in the HTC Vive Session, (5) Asynchronous Spacewarp (ASW) in the Oculus Rift session. Because NCSFI was not meant to be used in VR rendering, I had to make a few adaptations: to save on rendering time, only every other frame was rendered. These frames were used to create sharpened and blurry frames in accordance with the original design of the algorithm. For this comparison, I used the same blur method as for TRM, focusing only on the two fundamental differences between NCSFI and TRM: (1) NCSFI duplicates frames and (2) residuals are always added from sharp to blurry frames, regardless of motion. For ASW the content was rendered at 45 Hz and intermediate frames were generated using Oculus' implementation of ASW.

The computer-generated environments (Figure 5.9) consisted of an animated *football*, a *car* and *bedroom* (used only in Experiment 5.3). The first two scenes encouraged the observers to follow motion; the last one was designed to challenge screen-space warping. These scenes were rendered using the Unity game engine.

### Task:

Participants were asked to select the rendered sequence that had *better visual quality* and *motion quality*. Participants were presented with two techniques sequentially (10s each), with unlimited time afterwards to make their decisions. Before each session, participants were briefed about their task both verbally and in writing. For those participants who had never used a VR headset before, a short session was provided, where they could explore Valve's SteamVR lobby in order to familiarise themselves with the fully immersive environment. A pairwise comparison method with a full design was used, in which all

combinations of pairs were compared.

### Participants:

Nine paid participants aged 18–40 with normal or corrected-to-normal vision took part in Experiments 5.2 and 5.3. The majority of participants had little or no experience with virtual reality.

### Results:

The results of the pairwise comparison experiments were scaled using publicly available software<sup>2</sup> under Thurstone Model V assumptions in just-noticeable differences (JNDs), which quantify the relative quality differences between the techniques [Perez-Ortiz and Mantiuk 2017]. A difference of 1 JND means that 75% of the population can spot a difference between two conditions. The details of the scaling procedure can be found in [Perez-Ortiz and Mantiuk 2017]. Since JND values are relative, the 45 Hz condition was fixed at 1 JND for better presentation.

The results from Experiment 5.2 shown at top of Figure 5.10 indicate that the participants could not observe much difference between the proposed method and the original 90 Hz rendering. NCSFI improved slightly over the repeated frames (45 Hz) but was much worse than TRM or full-resolution rendering (90 Hz). Post-experiment interviews revealed that this could be due to ghosting (double edges) artefacts, which were well visible when blurred frames were displayed out of phase with misaligned residuals for fast head motion.

The results from Experiment 5.3 on the Oculus Rift, shown at the bottom of Figure 5.10, resemble the results of Experiment 5.2 on the HTC Vive: the participants could not observe much difference between TRM and a full 90 Hz rendering. ASW was seen to perform best in the *football* scene, whereas it performed worse in the *car* and *bedroom* scenes. This is because complex motion and colour variations in these scenes could not be compensated with screen-space warping, resulting in well visible artefacts. Note that although not included in this experiment, it is reasonable to expect that the technique by Didyk et al. [2010b] would suffer from similar warping artefacts, inherent to the use of re-projection.

### 5.5.2 Validation for desktop setup (Experiment 5.4)

The primary goal of this experiment was to compare the quality of TRM at three selected resolution reduction factors with standard rendering at 120 Hz and 60 Hz rendering. The setup was identical to the one used in Experiment 5.1 (2560 × 1440 G-Sync capable high-refresh-rate Asus ROG Swift P279Q 27" monitor, viewing distance fixed at 75cm using

---

<sup>2</sup>pwcmp - <https://github.com/mantiuk/pwcmp>

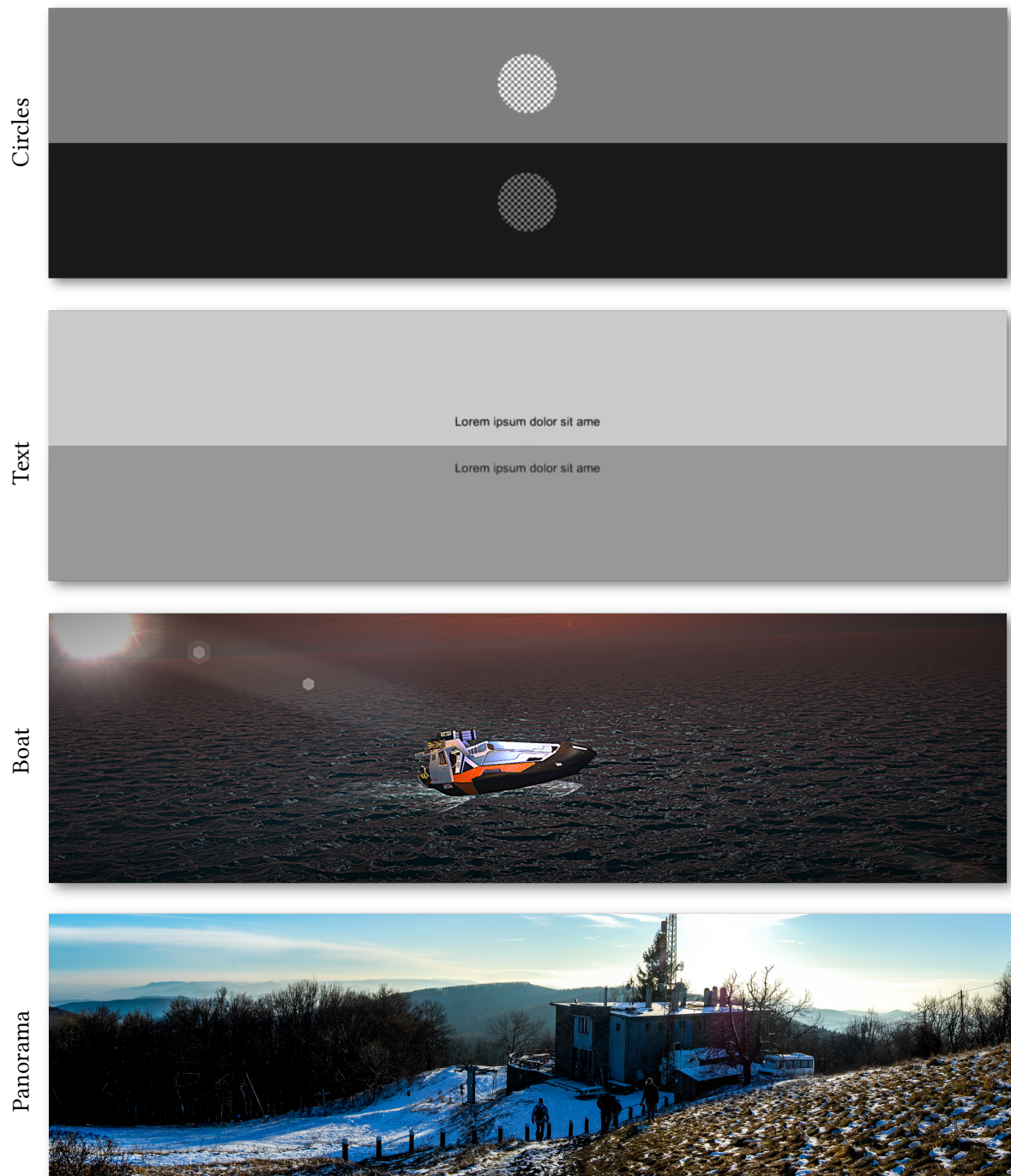


Figure 5.11: Stimuli used for validation on the high-refresh-rate monitor.

a headrest, Intel i7-7700, NVIDIA GeForce GTX 1080 Ti GPU). However, instead of comparing conditions sequentially, they were shown simultaneously side-by-side.

### **Stimuli:**

In each trial, the participant saw two 10-second looping video clips simultaneously, one in the upper, the other in the lower half of the screen. Five conditions were considered: (1) 60 Hz, presented at 120 Hz by repeating frames, (2) native 120 Hz video, (3–5) TRM with the odd-numbered frames reduced to  $\frac{1}{2}$ ,  $\frac{1}{4}$  or  $\frac{1}{8}$  of the original resolution. The clips *Circles*, *Text* and *Panorama* were identical to those used in Experiment 5.1, while the new clip *Boat* was added to test for more complex animation. The thumbnails of all clips are shown in Figure 5.11, while Figure 5.7 visualises how the eye perceives these videos when SPEM is taken into account. The clips were presented using custom software, which played uncompressed frames from the GPU memory.

### **Task:**

The task was identical as in Experiment 5.1, but the goal of the experiment was different — to measure the quality of each tested condition. A pairwise comparison method was used with the a full design, in which all combinations of pairs were compared. Each observer saw each pair three times, resulting in 120 trials per observer. The order of the stimuli as well as the position of the techniques on screen were randomised.

### **Participants:**

Eleven paid participants aged 18 – 40 took part in the experiment. All had normal or corrected-to-normal full colour vision.

### **Results:**

The results of the pairwise comparison experiments were scaled using publicly available software as in Experiments 5.2 and 5.3. Since JND values are relative, the 60 Hz condition was fixed at 1 JND for better presentation.

The results shown in Figure 5.12 indicate that observers could easily spot the difference between the 60 Hz and 120 Hz videos. TRM was nearly indistinguishable from 120 Hz for *Panorama* and *Text*, but about 75% of the observers could see the difference (1 JND) for *Boat* and *Circles* clips. This is consistent with findings in Experiment 5.1, only the threshold is shifted due to the side-by-side presentation. Further reduction in the resolution of odd frames did not result in a strong reduction of quality. Fortunately, as discussed, the saving in number of rendered or transmitted pixels also becomes negligible as the resolution is further reduced.



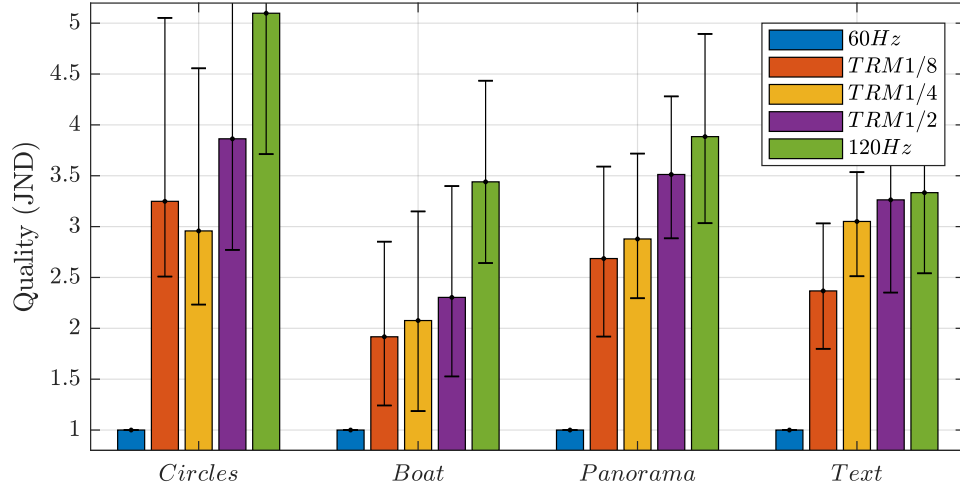


Figure 5.12: Results of experiment on a high-refresh-rate monitor. The higher JND values indicate higher quality. Error bars denote 95% confidence intervals.

## 5.6 Limitations

TRM is applicable only to high-refresh-rate displays, capable of showing 90 or more frames per second. At lower frame rates, flicker becomes visible. TRM is most beneficial when the angular velocities of movement are high, such as those introduced by camera motion in VR or first-person games. The proposed technique requires characterisation of the display on which it is used, as explained in Chapter 4. This is a relatively easy step for OLED-based VR headsets, but the characterisation is more involved for LCD panels.

Unlike re-projection techniques, TRM renders intermediate frames. This requires processing the full geometry of the scene, which might reduce performance gain for some scenarios. In particular, in scenes that are not fragment-bound, the main cost might come from physical simulation, or geometry processing. However, this cost is effectively amortised in VR stereo rendering, as explained in Section 5.4.1. The method also adds to the memory footprint as it requires additional buffers, for storing the previous frame and the residual. The memory footprint, however, is comparable to or smaller than that of re-projection methods.

TRM is rather sensitive to timing. The current pipeline does not offer a straightforward solution for recovering from dropped frames due to intermittent performance drops (e.g. operating system interrupt, poor cache conflict). Missing a frame will reduce the fundamental frequency at which TRM produces flicker, pushing it to a region where the human eye finds it highly objectionable. Future work should explore preventing such flicker potentially by falling back to a re-projection algorithm.

## 5.7 Summary

Temporal Resolution Multiplexing is an illustrative example of how a simple insight to the human visual system can guide the design of a rendering algorithm and consequently display design. In the coming years the visual quality of VR and AR systems is anticipated to improve primarily by increased display resolutions and higher refresh rates. However, rendering such a large number of pixels with minimum latency is challenging even for high-end graphics hardware. In this chapter I described an algorithm that reduces the GPU workload and the data sent to the display. TRM achieves a significant speed-up by requiring only every other frame to be rendered at full resolution. The method takes advantage of the limited ability of the visual system to perceive details of high spatial and temporal frequencies and renders a reduced number of pixels to produce smooth motion. TRM integrates easily into existing rasterisation pipelines, but could also be a natural fit with any fill-rate-bound high-frame-rate application, such as real-time ray tracing. A number of psychophysical experiments validated that TRM is close to being indistinguishable from full-resolution rendering while saving 42% of the computation power and data transfer bandwidth.

---

---

## CHAPTER 6

---

# MULTI-SCALE VISUAL MODELS

*“Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.”*

*George Box  
Science and Statistics*

We have seen how simple insights into the visual system can drive the design of more efficient graphics algorithms. I discussed a few examples in Chapters 2 and 3 such as chroma subsampling, and foveated rendering; and I argue that TRM, introduced in the previous chapter, is similar in spirit. Unfortunately evaluating and analysing such algorithms under novel conditions such as a wider colour gamut, higher dynamic range or novel display technology is non-trivial. Psychophysical evaluations require human participants which makes this approach impractical when dealing with a wide range of content or conditions.

Visual models can be used to gain a more robust prediction of the behaviour of human vision. Unfortunately, the HVS is remarkably complex, so a biologically accurate and content-wise robust end-to-end model is infeasible. Thus, a number of simplifying assumptions have to be made. A popular approach to reducing model complexity is to break up the visual stimulus into multiple visual bands, often corresponding to different scales of size.

In the next three chapters I describe the generic structure of such multi-scale models, and show how this can be applied to detecting a range of motion-related artefacts such as flicker (Chapter 7), judder and motion blur (Chapter 8).

This chapter borrows from the 2019 Human Vision and Electronic Imaging best-paper-award-winning article: *visual model for predicting chromatic banding artefacts*.

## 6.1 Multi-band models of vision

As discussed in Section 2.3, information throughout the visual system is processed by parallel pathways, and the final perception is often established by the aggregated output of these. The complex shape of the contrast sensitivity function can be also attributed to the multi-scale detector system of the visual system. Inspired by our knowledge of the visual system, the hard problem of detecting the visibility of a stimulus ( $I$ ) can be implemented by a collection of specialised detectors. The input of such models is the signal  $I$ , and the output is a joint detection probability, where this is typically interpreted as the probability of an average observer detecting the signal. State-of-the-art visual difference predictors, VDP and HDR-VDP, employ this idea when breaking up the difference signal between a reference image  $R$  and a distorted test image  $T$  into a number of spatial frequency bands one octave apart [Daly 2005; Mantiuk et al. 2005, 2011]. The model output is a map of probability values (indicating whether the difference can be detected up at a certain point of the image). The full design of HDR-VDP considers feature orientation bands, as well as inter-channel masking. In this chapter, I instead present a multi-channel model via the simplified problem of banding artefact detection.

## 6.2 Banding artefacts

To fit continuous colour values within a discrete, digital representation, linear colour values are quantised to a target bit-depth by transforming into some desired colour space. Quantisation of this sort can introduce banding artefacts, also known as quantisation artefacts or false contours. These banding artefacts are most prominent in images containing smooth, low texture regions, such as skin, or the sky and water in Figure 6.1. With insufficient bit-depth, smooth gradients in luminance and chrominance are perceived as wide, discrete bands.

A number of published works address the visibility and subsequent correction of banding artefacts [Lee et al. 2006b; Wang et al. 2016; Daly and Feng 2003, 2004]. Some authors consider the problem from an image processing point of view [Lee et al. 2006b] without a perceptual calibration. Wang et al. [Wang et al. 2016] demonstrate that such image processing methods can be fine-tuned to better correlate with subjective (mean opinion score) measurements, however, existing works take no account of the complex structure of visual perception, and are unlikely to generalize well to a colour space agnostic setup. Daly et al. [Daly and Feng 2003, 2004] present a technique for achieving bit-depth extension via spatio-temporal dithering. Their proposed technique utilizes the contrast sensitivity limitations of the human visual system to evaluate and recommend perceptually-ideal dithering patterns. In particular, [Daly and Feng 2004] analyses the error arising from

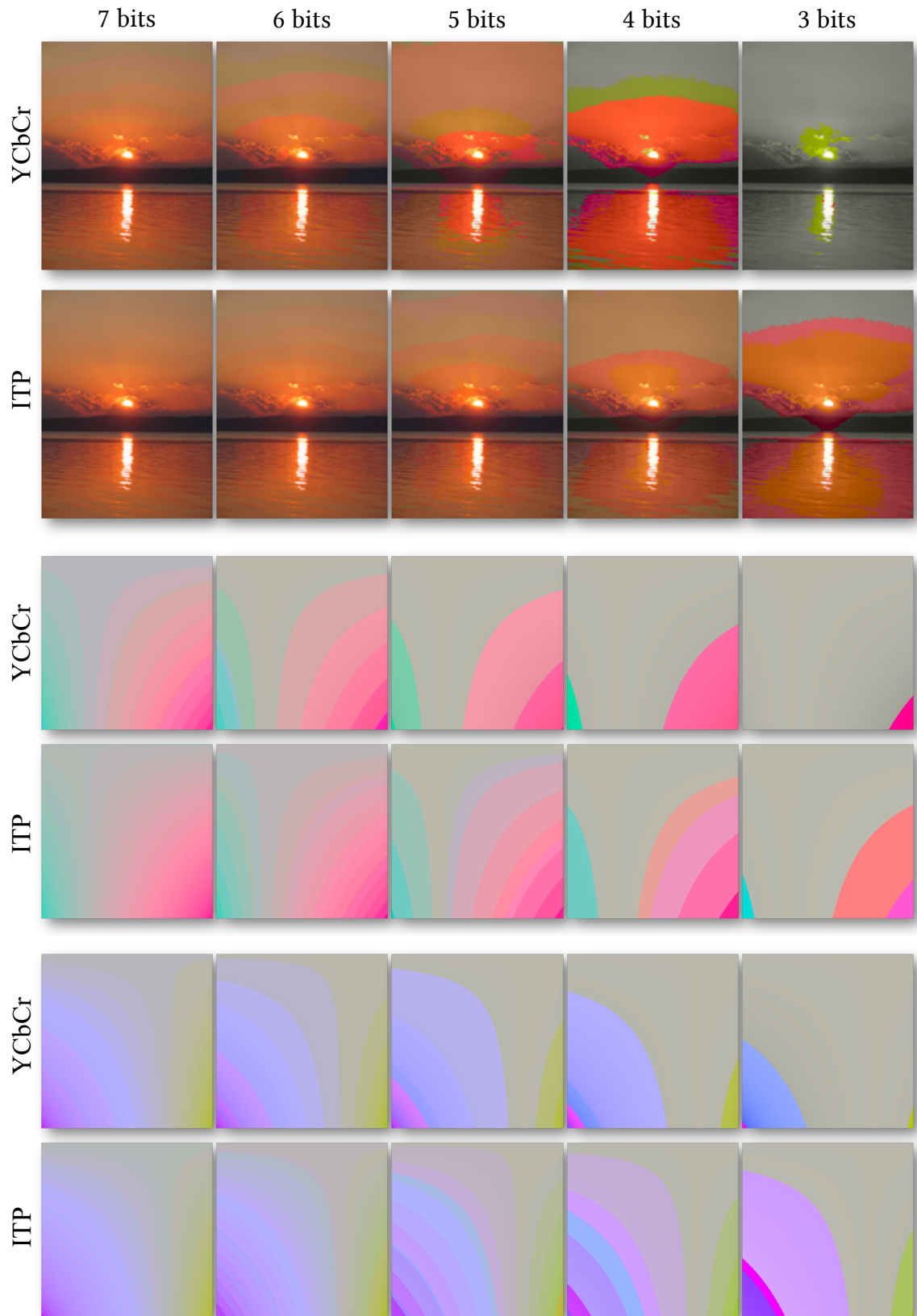


Figure 6.1: Increasing severity of banding artefacts when quantised to a range of bit-depths in different colour spaces (YCbCr, ITP). ITP results in less severe banding artefacts at the same bit-depth.

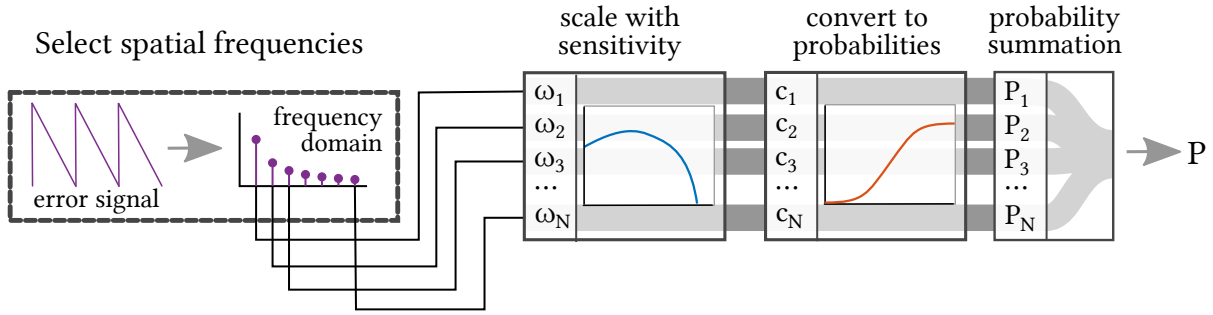


Figure 6.2: Overview of the banding artefact detection model for achromatic signals. The error signal is transformed into the frequency domain. The amplitude at each discrete frequency is scaled with the contrast sensitivity function, converted to probabilities with a psychometric function, then aggregated with probability summation to derive a single detection probability value.

the quantisation of a smooth gradient (Figure 6.3-left) in both the spatial and the frequency domains. The newly proposed visual model builds on this analysis and extends it for chromaticity.

## 6.3 Model design

As banding artefacts are the most visible when quantising smooth gradients, the presented model conservatively targets this the worst-case scenario. To break up the problem, I will first describe a model for achromatic images, then in Section 6.4 generalise to the chromatic model, demonstrating how the different colour channels can be aggregated. The design of this predictor model is following the typical pattern of multi-band models:

1. Determine the set of channels to operate on. In this case, we first determine a set of spatial frequencies that are present in the banding signal.
2. Scale each channel according to visual sensitivity. Here, we use the above-mentioned frequencies and Barten’s achromatic CSF model to determine how sensitive the eye is to each spatial frequency component.
3. Convert scaled sensitivities to detection probabilities. This is typically done with a psychophysical function such as the cumulative Weibull distribution, or the cumulative normal distribution.
4. Pool probabilities across each channel. Similarly to VDP and HDR-VDP, I use probability summation.

The rest of this section explains the above steps in more detail. For a schematic diagram of the achromatic architecture, see Figure 6.2.

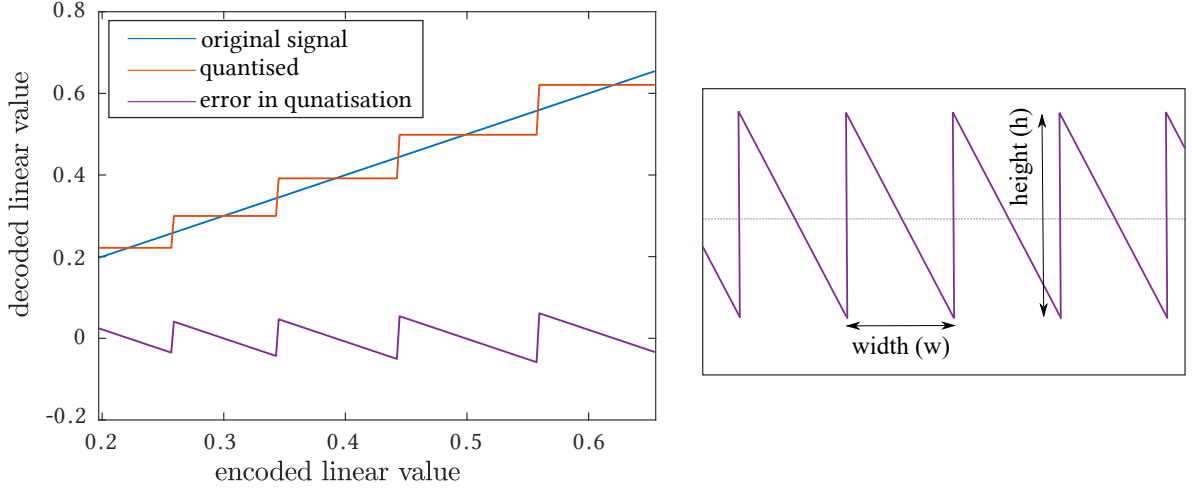


Figure 6.3: Left: Illustration of the error signal (purple) between the original (blue) and the quantised (red) signals. The error signal approximates a saw-tooth function. Right: Analytical model of the error signal. With a known width ( $w$ ) and height ( $h$ ) we can find the frequency components of the signal.

## Identifying the spatial frequencies

To determine spatial frequencies of the contours, we can analyse the Fourier transform of the difference signal between the smooth and contoured gradients. The banding artefacts, *i.e.* the difference between smooth and quantised signals over a smooth gradient can be approximated with a saw-tooth pattern (see purple line in Figure 6.3):

$$f(x) = w \lfloor x/w \rfloor - x, \quad (6.1)$$

where  $x$  goes from 0 to 1, and  $w$  is the number of quantisation levels. The Fourier series is given by:

$$f(x) = \frac{-h}{2} + \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{h}{k} \sin \left( \frac{k2\pi x}{f} \right), \quad (6.2)$$

where  $x$  is spatial location in visual degrees,  $w$  is the width (period) and  $h$  is the amplitude as in Figure 6.3. The amplitude of the frequency components is then:

$$\alpha_k = \frac{h}{k\pi} \quad \text{for } k = 1, 2, \dots \quad (6.3)$$

and the frequency of each component is:

$$\omega_k = \frac{k\rho}{w}, \quad (6.4)$$

where  $\rho$  denotes the angular resolution of the device in pixels per visual degree and  $w$  is the width of the saw-tooth in pixels. As the eye has very limited sensitivity to high

spatial frequencies, in my findings Fourier components for  $k > 16$  were insignificant and did not improve the model's accuracy.

## Scaling with sensitivity

To compute the probability of detecting each Fourier component of the contour, we can first determine the sensitivity to that component :

$$S = \frac{L_b}{\Delta L_{det}} = \rho_A(\omega, L_b), \quad (6.5)$$

where  $\omega$  is the spatial frequency,  $L_b$  is background luminance,  $\Delta L_{det}$  is the detectable amplitude of that frequency component, and  $\rho_A()$  is Bartens' achromatic CSF model. Then, I normalise the contrast of the contouring pattern ( $a_k/L_b$ ) by multiplying by the sensitivity so that the normalised values are equal to 1 when the  $k$ -th frequency component is just detectable. The normalised contrast is given by:

$$c_k = \frac{a_k}{L_b} \rho_A(\omega_k, L_b). \quad (6.6)$$

## Conversion to probabilities

Next, to transform the normalised contrast into probabilities, use the Weibull psychometric function [Weibull 1951]:

$$P_k = 1 - \exp(\ln(0.5)c_k^\beta), \quad (6.7)$$

where  $\beta$  is the slope of the psychometric function. An estimate of  $\beta = 3.5$  is common in visual sciences, however, for this particular model,  $\beta = 2$  produced more uniform outputs (see Figure 6.5).

## Aggregating probabilities

Finally, the probabilities  $P_k$  need to be pooled across all Fourier components. In multi-band models, where each band represents independent detection probabilities, probability summation (PS) is normally used. Note that “probability summation” can be considered a misnomer [Baldwin et al. 2015], as it represents the inverse of the probability of none of the channels detecting the difference. No summation is involved.

$$P = 1 - \prod_k (1 - P_k) \quad (6.8)$$



To determine the minimum bit-depth that does not result in contouring artefacts, one can then perform binary-search for a bit-depth that would result in  $P = 0.5$ .

## 6.4 Extension for chromatic banding

The model, as described above, accounts only for banding due to changes in luminance. While change in luminance is usually a large contributor to the visibility of banding, changes in chromaticity can also have an impact. We can extend the model to account for this using a chromatic contrast sensitivity function, and employ probability summation across all visual channels. The final chromatic colour discrimination model takes a colour gradient, specified in the CIE XYZ (1931) colour space, and predicts the probability of observing a banding artefact when the smooth gradient is quantised to a given bit-depth.

First, independence of the visual channels is a crucial assumption, as this model does not take masking into account. Hence, I convert all colours from  $XYZ$  into  $LMS$  space (assuming CIE 1931 colour matching functions). Each channel of this tri-stimulus space is proportional to the response of the long, medium and short cones of the retina (see Section 2.2). It should be noted that there is no standard way to scale the absolute response of each cone type and the response values are only relative. To convert CIE  $XYZ$  trichromatic values into  $LMS$  responses I use the following linear transform:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.15514 & 0.54312 & -0.03286 \\ -0.15514 & 0.45684 & 0.03286 \\ 0 & 0 & -0.00801 \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \quad (6.9)$$

The cone responses are further transformed into opponent responses: one achromatic (black-to-white) and two chromatic: red-to-green and yellow-green-to-violet. The exact tuning colour directions of those mechanisms are unknown, so I use one of the simplest formulae commonly used in the literature:

$$\begin{bmatrix} A \\ R \\ Y \end{bmatrix} = \begin{bmatrix} L + M \\ L - M \\ S \end{bmatrix}, \quad (6.10)$$

where  $A$  is achromatic (luminance) response,  $R$  is the red-to-green response and  $Y$  is the yellow-green-to-violet response.

Given two colours to be discriminated, we need to compute contrast between them. Since there is no single way to represent colour contrast, my collaborators and I experi-

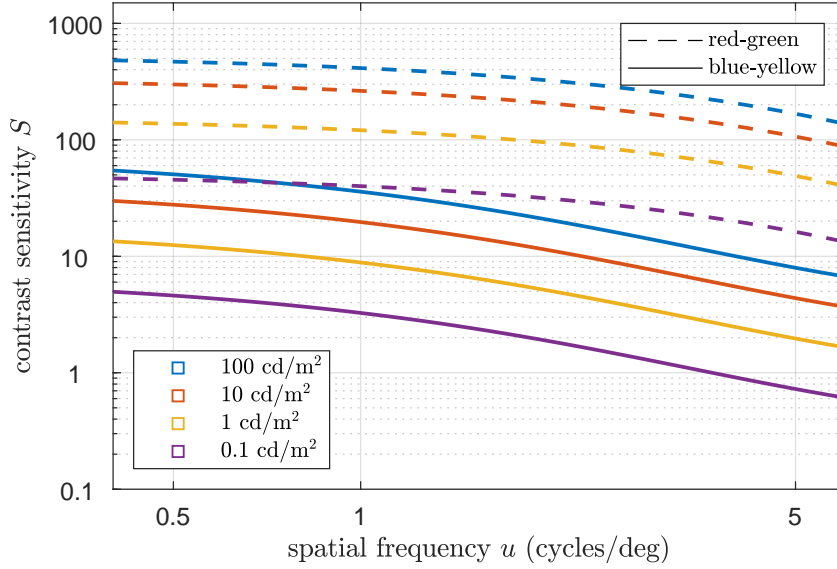


Figure 6.4: Chromatic contrast sensitivity function based on [Kim et al. 2013]. Dashed and solid lines show the sensitivity to red-green and blue-yellow change respectively at different background luminance levels.

mented with a number of expressions to find the most suitable for this model:

$$C_A = \frac{|A_2 - A_1|}{A_1}, C_R = \frac{|R_2 - R_1|}{\alpha|R_1| + (1 - \alpha)A_1}, C_Y = \frac{|Y_2 - Y_1|}{\alpha|Y_1| + (1 - \alpha)A_1}, \quad (6.11)$$

where  $\alpha$  is a free variable to be optimised for the experiment data. Given the colour contrast components  $C_A, C_R, C_Y$ , we follow the same steps as for the prediction of luminance banding: (1) each colour contrast is multiplied by the corresponding contrast sensitivity function from [Kim et al. 2013] and the Fourier coefficients of the saw-tooth pattern,  $a_k$ , (2) then transformed to detection probability, (3) then apply probability summation across all frequencies and the  $A, R, Y$  colour channels.

## 6.5 Model predictions

The model can be trivially extended to take a starting colour and a colour direction vector as input (instead of the smooth image gradient). Binary search can then establish the colour along the colour direction vector for which the probability of detectable banding artefacts is 0.5. I use such extension of the model to establish a detection threshold and to plot colour uniformity ellipses akin to MacAdam ellipses. Figure 6.5 compares the predictions of the proposed model to CIE  $\Delta E$  2000 difference and to the original MacAdam ellipses. Note that the proposed model is meant to provide better predictions for banding rather than predicting traditional colour patch difference; hence it is an interesting observation that the resulting shapes are comparable.

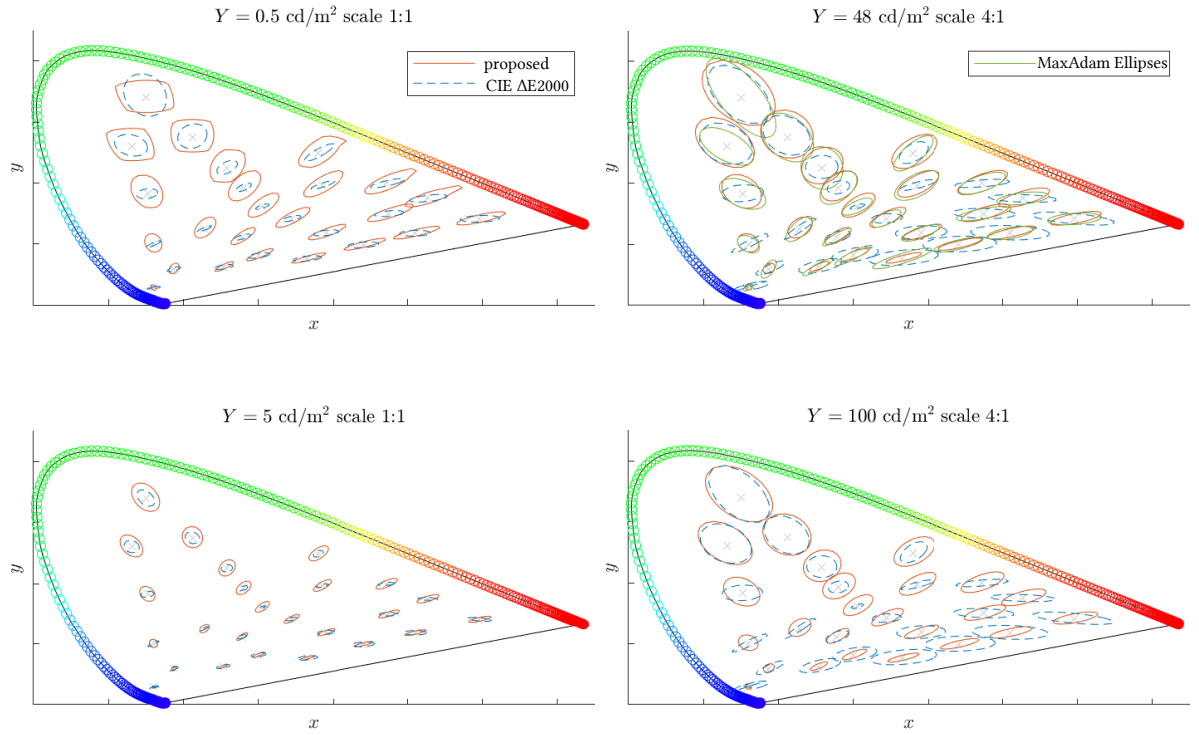


Figure 6.5: Predictions of the proposed model compared to CIE *DeltaE* 2000 and MacAdam ellipses [Brown 1957]. Each plot corresponds to different luminance level. Note that MacAdam ellipses were measured only for the background luminance of  $48 \text{ cd/m}^2$

## 6.6 Summary

Simple insights into the visual system are powerful. However, to create, calibrate and analyse novel graphical algorithms, we need to either rely on a new psychophysical experiment every time, or a visual metric that is known to correlate with subjective preference. The human visual system is extremely complex, so a number of simplifying assumptions have to be made. One such assumption is the presence of multiple independent channels which together contribute to the final perception. In this chapter, I introduced the concept of multi-band models via the problem of chromatic banding artefact detection.

Such white-box models have been demonstrated to produce accurate predictions only after the model parameters (parameters of the Barten model in this case) are adjusted. For more detail for how this was achieved for banding detection, please refer to the full paper [Denes et al. 2019a].

---

---

# CHAPTER 7

---

## A VISUAL MODEL FOR FLICKER

Novel display algorithms, such as 1 black frame insertion, and temporal resolution multiplexing (Chapter 5) introduce temporal change into images at 40-180 Hz, on the boundary of the temporal integration of the visual system. This can lead to flicker, which in turn induces viewer discomfort. The critical flicker frequency (CFF) alone does not model this phenomenon well, as flicker sensitivity varies with contrast, and spatial frequency; a content-aware model is required. In this section, I introduce a visual model for predicting flicker visibility in temporally changing images. The model performs a multi-scale analysis on the difference between consecutive frames, normalising values with the spatio-temporal contrast sensitivity function as approximated by the pyramid of visibility. The output of the model is a 2D detection probability map. I also describe the subjective flicker marking experiment I ran to fit the model parameters, then analyse the difference between two display algorithms, black frame insertion and temporal resolution multiplexing, to demonstrate the application of the model. This chapter borrows from the 2020 Human Vision and Electronic Imaging article: *Predicting visible flicker in temporally changing images*.

### 7.1 Model design

Temporal multiplexing algorithms often manipulate pairs of frames. Let us denote two consecutive frames as  $F_i$  and  $F_{i+1}$  that could be, for instance, the reduced-resolution and sharpened frames of TRM; or a black frame and a luminance-boosted frame of BFI. The proposed visual model predicts whether displaying  $F_i$  and  $F_{i+1}$  alternately at refresh rate  $R$  would result in perceivable flicker. The output is a  $P_{\text{det}}(x, y)$  map corresponding to the percentage of the population detecting flicker at a pixel  $(x, y)$ .

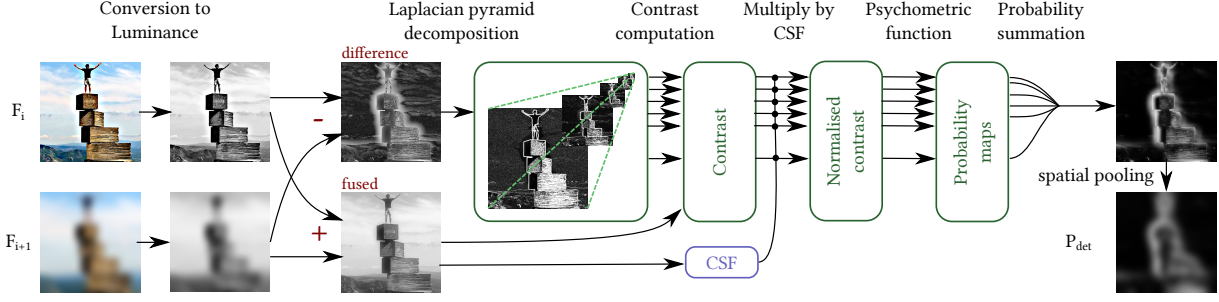


Figure 7.1: Overview of the flicker predictor model. The input is a pair of colour frames; the result of the model is a 2D probability of detection ( $P_{\text{det}}$ ) map.

The flicker predictor utilises a spatio-temporal CSF in a multi-scale model with probability summation along the spatial frequency bands. For an overview of the pipeline, please see Figure 7.1. The model does not distinguish orientation-sensitive bands, often found in masking models [Watson and Solomon 1997]. The model also assumes that eye movements have been already accounted for in image-space; *i.e.*, the same pixels on  $F_i$  and  $F_{i+1}$  will correspond to roughly the same photoreceptors on the retina.

As the source of most flicker is the change in luminance, I do not consider chromaticity here. First, respective luminance values are computed ( $Y_i$  and  $Y_{i+1}$ ) of consecutive frames ( $F_i$  and  $F_{i+1}$ ) based on a calibrated display model. Then, to find contrast, compute the difference image:

$$\Delta(x, y) = Y_i(x, y) - Y_{i+1}(x, y), \quad (7.1)$$

where  $x$  and  $y$  describe pixel location, and  $Y_i(x, y)$  is the luminance of pixel  $(x, y)$  of the frame  $F_i$ . The summed luminance of the consecutive frames can be similarly defined as:

$$\bar{Y}(x, y) = Y_i(x, y) + Y_{i+1}(x, y) \quad (7.2)$$

## Spatial frequency decomposition

As flicker sensitivity varies with spatial frequency, the difference image  $\Delta(x, y)$  is decomposed into a Laplacian pyramid. This, in essence, creates the multiple bands of the visual model. Each layer of the pyramid is half the spatial resolution of the one above; the bottom layer capturing 2 cycles per visual degree (cpd) resolution or just below – e.g. for a 52 pixel-per-degree image the mid points of the spatial frequency bands are  $S_i = \{26, 13, 6.5, 3.25, 1.625\}$  cpd. I use an undecimated pyramid, in which each band has the same resolution.

## Contrast scaling with sensitivity

In each layer, Michelson contrast can be then computed as:

$$C(x, y, l) = \frac{|\Delta(x, y, l)|}{\bar{Y}(x, y)}, \quad (7.3)$$

where  $l$  is the Laplacian pyramid layer. To account for contrast sensitivity, contrast is normalised at each layer by a the spatio-temporal CSF ( $\rho$ ). I use the parametric pyramid of visibility, as it has been shown to provide a good fit to previous CSF measurements [Watson and Ahumada 2017].

$$\rho(W, F, Y) = \exp(c_0 + c_W W + c_F F + c_L \log Y), \quad (7.4)$$

where  $W$  and  $F$  are the spatial and temporal frequencies in cpd and Hz, as in the original paper,  $Y$  is the adapting luminance in  $\text{cd/m}^2$ , and  $(c_0, c_W, c_F, c_L)$  are parameters that are kept as free variables in the model. The underlying assumption is that the mean fused image  $0.5\bar{Y}(x, y)$  provides a good estimate of the local adapting luminance. The normalised contrast is then:

$$\hat{C}(x, y, l) = C(x, y, l)\rho(R/2, S_l, 0.5\bar{Y}(x, y)), \quad (7.5)$$

where  $S_l$  is the spatial frequency of the layer, and  $R$  is the display refresh rate. I use  $R/2$  to sample the temporal dimension of the CSF, as when modifying pairs of frames, this is the highest temporal frequency according to the Nyquist limit.

## Conversion to probabilities

Next, to transform the normalised contrast into probabilities of detection, a Weibull psychometric function is used:

$$P(x, y, l) = 1 - \frac{\exp(\hat{C}^\beta(x, y, l))}{2}, \quad (7.6)$$

where  $\beta$  controls the slope of the psychometric function, a free parameter. In order to pool the probabilities across all layers, probability summation is used to compute  $P(x, y)$ .

Finally, to account for detection inaccuracies, the map is further convolved with a small Gaussian filter. Doing this in probability space yielded surprisingly better predictions than in contrast space:

$$P_{\text{det}}(x, y) = P(x, y) * G_{\sigma_{sp}}, \quad (7.7)$$

where  $*$  denotes convolution, and  $G_{\sigma_{sp}}$  is a Gaussian kernel with  $\sigma_{sp}$  being a free parameter in the model.



Figure 7.2: Pool of reference images used for the flicker marking experiment with a range of content.

## 7.2 Psychophysical calibration (Experiment 7.1)

To tune the model parameters, ground-truth data is required on flicker perception in complex images. As flicker is often perceived in multiple parts of the image, and location information is crucial for  $P_{\text{det}}(x, y)$ , a marking experiment was chosen. Such experiments have been utilised to calibrate similar metrics for image difference predictors [Wolski et al. 2018; Ye et al. 2019].

### Participants

Nineteen participants aged 18-40 with normal or corrected-to-normal vision took part in the experiment. As the trials involved looking at flickering images, a pre-experiment screening question was used to ensure that no participant reported a history of epilepsy. Informed consent was acquired before the beginning of the experiment, which involved briefing participants on the aim, the procedure, and potential risks of the experiment both verbally and in writing.

### Setup

Participants were shown a  $512 \times 512$  pixel flickering photograph in the centre of the ASUS monitor. The viewing distance was fixed at 65 cm, yielding an angular resolution of 52 pixels per degree (ppd). Images hence had a field of view of  $9.85^\circ$ ; the rest of the monitor was filled with a grey background of  $36 \text{ cd/m}^2$ . Accurate refresh rates were achieved with custom C++/OpenGL software and NVidia G-Sync.

### Stimuli

Eighteen stimuli were created by flickering twelve colour photographs (see Figure 7.2). The photographs provided a range of content from primitive zebra stripes, photographs of



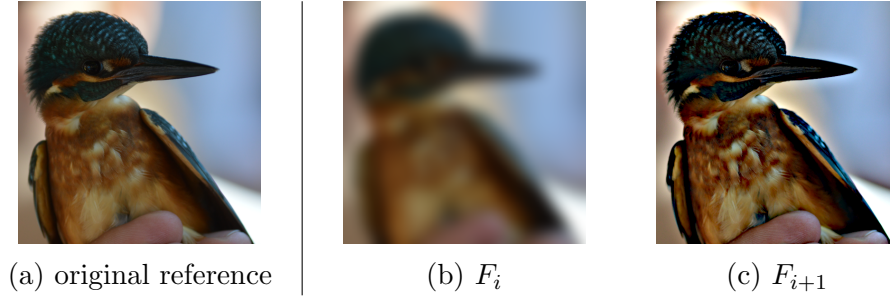


Figure 7.3: Example stimulus pair. Similarly to TRM, reference frame is low-pass filtered ( $F_i$ ), and sharpened ( $F_{i+1}$ ) to produce band-limited flicker.

birds, buildings and people. For each trial, a spatially band-limited flicker was introduced at temporal frequency  $R$ . For each trial, a pair of frames were computed ( $F_i$  and  $F_{i+1}$ ) and alternated at refresh rate  $R$  in a manner similar to TRM:

$$\begin{aligned} F_i(x, y) &= F_{\text{ref}}(x, y) * g_{\sigma}(x, y) \\ F_{i+1}(x, y) &= \xi^{-1}\left(2\xi(F_{\text{ref}}(x, y)) - \xi(F_i(x, y))\right), \end{aligned} \quad (7.8)$$

where  $F_{\text{ref}}$  denotes the original reference image in gamma-compressed *rgb* colour space (rec.709 primaries),  $*$  stands for 2D convolution,  $g_{\sigma}(x, y)$  is an isotropic Gaussian blur kernel with a standard deviation of  $\sigma$ , and  $\xi()$  is a gain-offset-gamma display model [Berns 1996] transforming gamma-compressed *rgb* to linear *rgb* values. For stimulus generation, an approximation of the ASUS display model was used:

$$r'(x, y) = 0.99917r(x, y)^{2.15} + 0.000825, \quad (7.9)$$

where  $r$  is the red channel. The same formula was applied to all colour channels. For an example image pair, see Figure 7.3. Note that unlike in TRM, the residual buffer is not used here. This might introduce some spatial artefacts to the viewer, but this experiment did not attempt to establish overall visual quality, and such artefacts did not impede flicker detection. For a summary of the stimuli images,  $\sigma$  and  $R$  values, refer to Figure 7.4. For a complete illustration, please see Appendix B.

## Task

Participants were asked to “mark (or *paint*) any part of the image where flicker is visible” – quoted from the briefing form. Flickering areas could be marked by holding down the left mouse button and moving the pointer around. Previous markings could be deleted with the right mouse button in a similar fashion. A circular mouse pointer was used with the diameter adjustable from  $0.15^\circ$  (8 pixels) to  $2^\circ$  (104 pixels) using the mouse wheel.

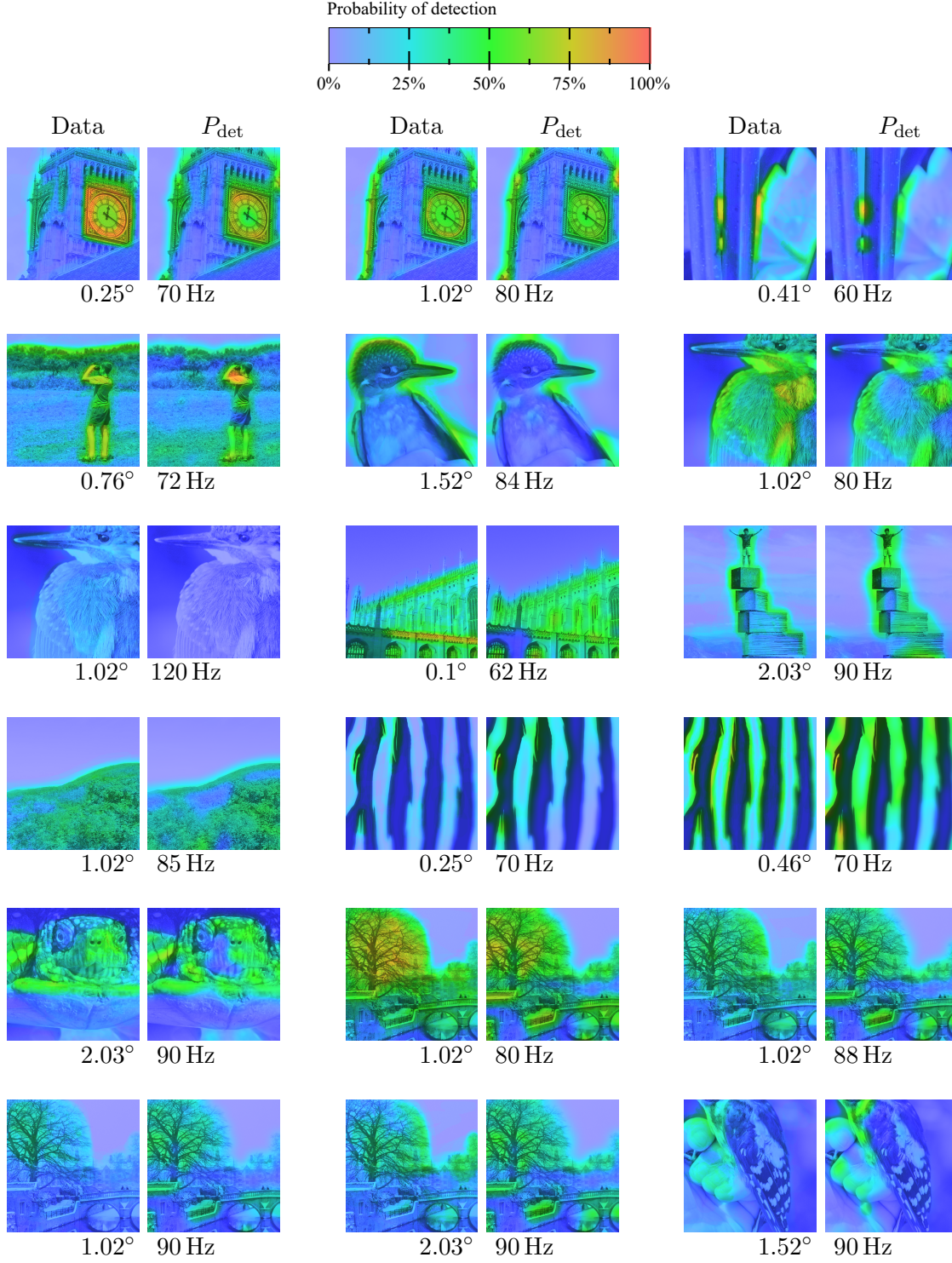


Figure 7.4: Flicker markings (Data) and model predictions ( $P_{\text{det}}$ ) overlaid on the 18 reference images. Blue indicates no flicker ( $P_{\text{det}} \approx 0$ ), green to orange indicates strong perceivable flicker ( $P_{\text{det}} \rightarrow 1$ ). Sub-captions state the standard deviation of the Gaussian kernel (in visual degrees), and the refresh rate at which  $F_i$  and  $F_{i+1}$  were alternated (Hz).

Any marked area was highlighted and immediately stopped flickering. At the beginning of each trial, the mask was cleared. Participants were specifically asked to first mark (and hence remove) the strongest flicker first to minimise the effect of masking. During briefing and training it was highlighted that flicker might be more visible in the parafoveal area of vision, and hence looking at objects slightly off-centre might reveal more flicker. This was to ensure that all participants utilise the free-viewing setup equally.

Each participant created a marking map for each stimulus three times, yielding 54 trials in total. The order of the trials was randomised.

## Results

Figure 7.4 shows the flicker marking maps averaged over nineteen observers and three repetitions. As expected, flicker perception degrades with increasing refresh rates and increases with the blur  $\sigma$ . The markings, however, cannot be considered ground truth data for two reasons: (1) participants might make mistakes producing mis-markings, and (2) the finite size of the brush allows for limited precision.

Markings can be considered the output of a stochastic process, where observers attend to a distortion with  $P_{\text{att}}$ , and mis-mark a pixel with probability  $P_{\text{mis}} = 0.01$  [Wolski et al. 2018]. Due to the small image size ( $9.85^\circ \times 9.85^\circ$ ), and the characteristics of temporal sensitivity, I assumed  $P_{\text{att}} = 1$  for this experiment.

For each image in each 57 marking maps, each  $(x, y)$  pixel takes a binary  $\{0, 1\}$  value depending on whether the participant marked it with the mouse. Assuming a detection probability  $P_{\text{det}}(x, y)$ , the data can be modelled as a binomial distribution. Accounting for the mis-markings, the likelihood of observing the collected data given a model is:

$$\Lambda(x, y) = P_{\text{mis}} + (1 - P_{\text{mis}}) \binom{n}{k} P_{\text{det}}(x, y)^k [1 - P_{\text{det}}(x, y)]^{n-k}, \quad (7.10)$$

where  $n = 57$  is the number of all collected markings for an image,  $k$  is the number of trials a pixel is marked as flickering, and  $P_{\text{det}}(x, y)$  is the predicted detection probability.

## Parameter fitting

The task of finding the best model parameters can be posed as a non-linear optimisation problem, maximising the average log-likelihood over the images. However, I observed that the effects of spatial pooling were masked by the finite paint brush size; therefore decided to fix this parameter to a value comparable to the brush sizes ( $\sigma = 0.36^\circ$ ). Variable slope values in the psychometric function were also expected to create a range of local minima, hence I selected a single likely candidate  $\beta = 2$ . The remaining parameters are parameters from the pyramid of visibility which were restricted to physically sensible ranges ( $c_W < 0$  for decreasing sensitivity with temporal frequency;  $c_F < 0$  for decreasing sensitivity with

	<b>training</b>	<b>test</b>	$c_0$	$c_W$	$c_F$	$c_L$
Pyramid of visibility	-	-3.204	2.1900	-0.0600	-0.0650	0.3880
Fit to all	-2.849	-	1.9993	-0.1059	-0.0242	0.9102
cross 1	-2.829	-2.933	2.1946	-0.1025	-0.0362	0.8581
cross 2	-2.903	-2.779	1.9982	-0.1146	-0.0113	0.9782
cross 3	-2.782	-3.009	1.5741	-0.1006	-0.0128	0.9265

Table 7.1: Parameters, training and test fitness (measured as mean log-likelihood). Pyramid of visibility (first row) uses the parameters from the Robson fit from [Watson and Ahumada 2017]. When fitting to the new dataset, the log likelihood increases (Fit to all). Cross # are indicate results of the 3-fold cross-validation.

spatial frequency;  $c_L > 0$  for increasing sensitivity with background luminance).

The results are summarised in Table 7.1. When refitting the parameters from the original pyramid of visibility, the mean log-likelihood increases, as expected. Parameters show some deviation from the values fitted to the Robson measurements. While  $c_0$  is comparable, increasing temporal frequencies attenuate sensitivity faster (lower  $c_W$  values), increasing spatial frequencies attenuate sensitivity slower, and luminance amplifies sensitivity faster. Such deviations are to be expected due to the significantly more complex nature of the task presented in the marking experiment.

To analyse the possibility of over-fitting the model parameters to the new dataset, a 3-fold cross-validation was executed. For this, the dataset was randomly split into three 6-element groups. The models was fit to each two groups (training), then performance was evaluated on the third groups (test). Results in Table 7.1 indicate that the training and test likelihoods were comparable, and the optimum parameters did not differ significantly from the scenario when all 18 images were included in the training dataset.

Qualitatively, I observed that model predictions for the experiment stimuli capture the flickering details well. User-produced and predicted markings are as shown in Figure 7.4.

## 7.3 Application

The proposed flicker model can be used to analyse novel temporal multiplexing algorithms. To demonstrate this, I analyse the amount of flicker introduced by TRM and BFI. For TRM I assume the worst-case scenario and ignore the residual buffer. For black frame insertion  $F_i$  was set completely black, while  $F_{i+1}$  was boosted to double the luminance. For representative content, three computer-generated images were selected.

As shown in Table 7.2, TRM generally requires lower refresh rates; it is unlikely to be perceived as flickery on 90 Hz, while BFI causes minor distortions even on 120 Hz. This is consistent with previous observations Chapter 5, specifically the claims outlined in Table 5.1.

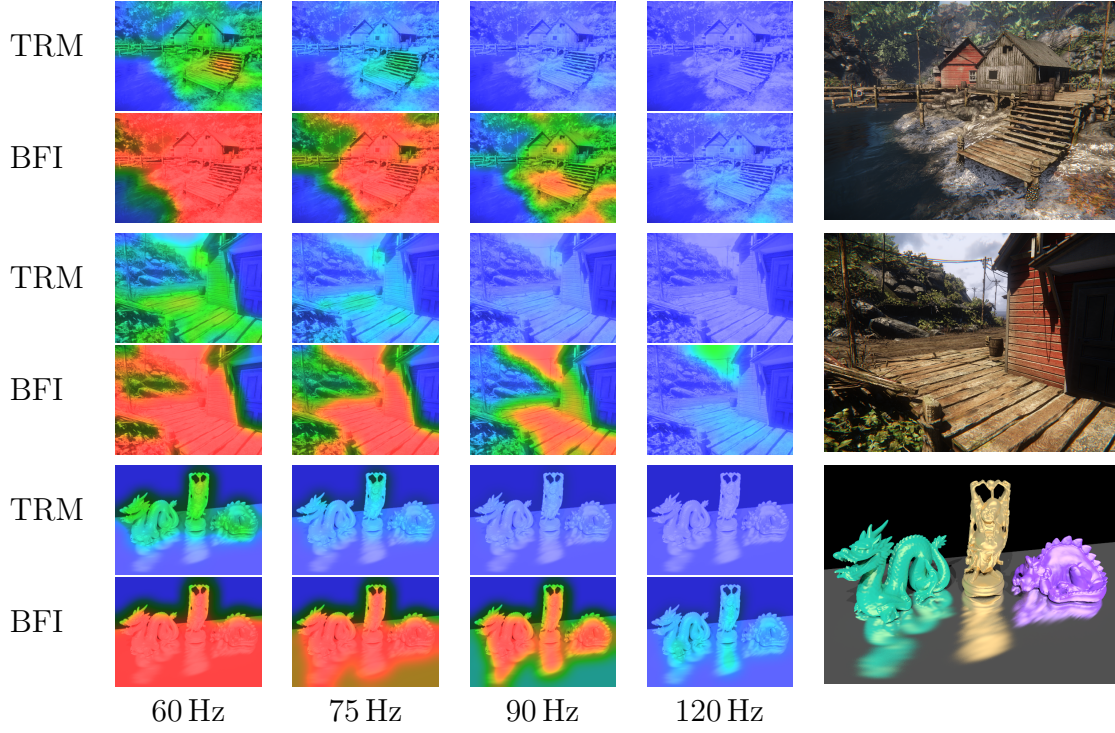


Table 7.2: Analysis of flicker artefacts in black frame insertion (BFI) and TRM across refresh rates. Red regions indicate that BFI produces noticeable flicker at bright regions even on 120 Hz. TRM, on the other hand, does not flicker above 90 Hz on the analyzed content for the target display with  $156 \text{ cd/m}^2$  peak luminance.

## 7.4 Summary

Novel display algorithms introduce temporal change which can lead to flicker, which in turn induces viewer discomfort. The critical flicker frequency (CFF) alone does not model this phenomenon well, as flicker sensitivity varies with contrast, and spatial frequency. In this chapter, I introduced a content-aware visual model for predicting flicker visibility in temporally changing images. The model performs a multi-scale analysis on the difference between consecutive frames, normalising values with the spatio-temporal contrast sensitivity function as approximated by the pyramid of visibility. The output of the model is a 2D detection probability map. I described the subjective flicker marking experiment I ran to fit the model parameters. Finally, I demonstrated how the new model can be used to analyse the difference between two display algorithms, black frame insertion and temporal resolution multiplexing.

Flicker is one of the four motion artefacts (Section 2.6.3). In the next chapter, I introduce a visual model for detecting two other motion artefacts (judder and motion blur), and based on the model predictions, I propose a novel motion-adaptive rendering algorithm.



---

## CHAPTER 8

---

# VISUAL MODEL FOR BLUR AND JUDDER

Avoiding flicker artefacts is crucial for temporal multiplexing techniques. However, in traditional rendering, blur and judder artefacts are more prominent. The perceived effect of such motion artefacts, especially in the range of high-refresh-rate displays ( $> 60$  Hz) is not well-understood. In this section, I describe psychophysical measurements of perceived visual quality of motion from 50 Hz to 165 Hz. Then, I propose a novel visual model that predicts the quality taking into account two motion artefacts: non-smooth motion (judder) and blur. Unlike similar work, I incorporate the velocity and predictability of motion into the model. The treatment of blur is also uniquely broad, fusing hold-type blur, eye motion and spatial blur arising from resolution reduction. The model isolates individual components contributing to the quality of motion, such as judder, spatial blur aligned with the direction of motion, and spatial blur orthogonal to the direction of motion. Then, the discrimination of each component is modelled using spatio-temporal contrast sensitivity functions. To find the free parameters of the model, I measured eye movement accuracy and conducted further psychophysical experiments to quantify motion quality.

First, I describe the psychophysical experiment for measuring motion quality, then introduce the model and describe the additional experiments used to fit the free parameters in the model. Finally, at the end of the chapter, I propose a novel motion-adaptive refresh rate and resolution rendering algorithm alongside a psychophysical validation experiment.

This chapter is based on the SIGGRAPH 2020 article titled *A perceptual model of motion quality for rendering with adaptive refresh-rate and resolution*.



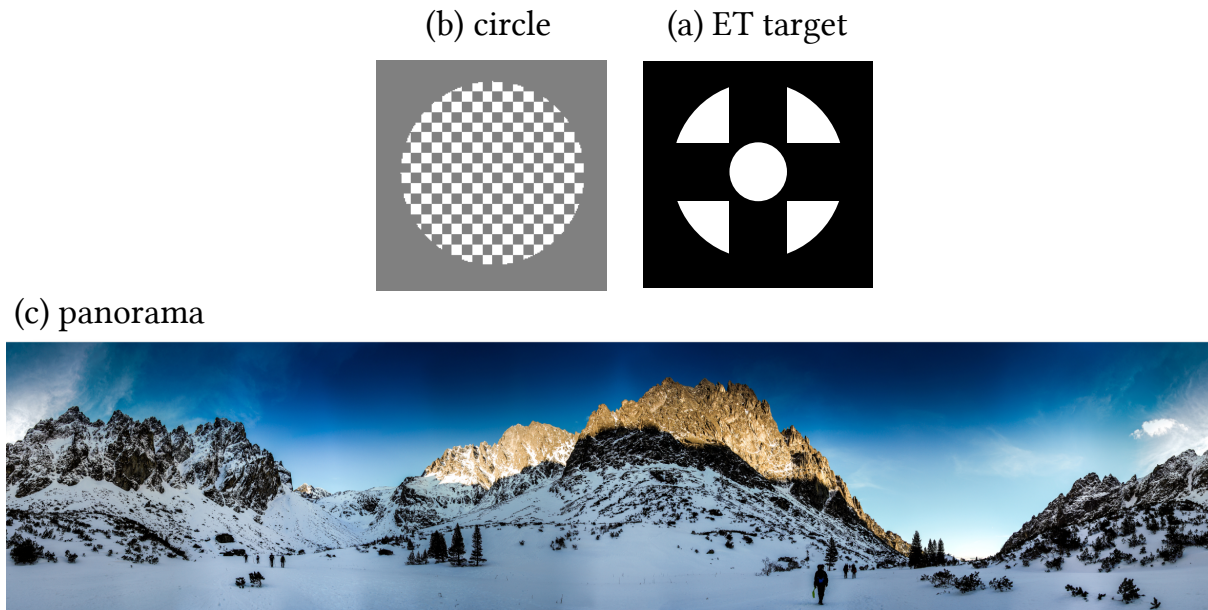


Figure 8.1: Animations used in the motion quality experiment.

## 8.1 Measuring motion quality (Experiment 8.1)

To motivate the design of the visual model, I first measured how perceived visual quality changes with refresh rate, motion velocity and the predictability of motion. A pairwise comparison protocol was chosen because of its relative simplicity and speed [Zerman et al. 2018], converting to a one-dimensional scale using Thurstone case V assumptions as in TRM. To efficiently measure quality across the uniquely wide range of conditions, I relied on an active sampling technique developed by a co-author.

### Setup

Observers were shown the same scene at two different refresh rates, each shown on a separate G-sync capable ASUS display. The displays were stacked on top of each other to make the task of comparing horizontal motion easier. The viewing distance was 108 cm (30° field of view). Every animation was shown at a one of 23 refresh rates from 50 Hz to 165 Hz. The granularity of 5 Hz was chosen to approximate the 1 JND threshold at 50 Hz [DoVale 2017].

### Stimuli

To cover a range of realistic and synthetic content, three animations were used: checkered circle (*circle*), eye tracker target (*ET*), and a panorama image (*panorama*) (Figure 8.1). *ET* was a combination of a bull’s eye and a cross hair which has been shown to be effective as a fixation target [Thaler et al. 2013]. Animations had only horizontal motion to aid comparison in the vertically-stacked setup. Each of 6 tested conditions involved different



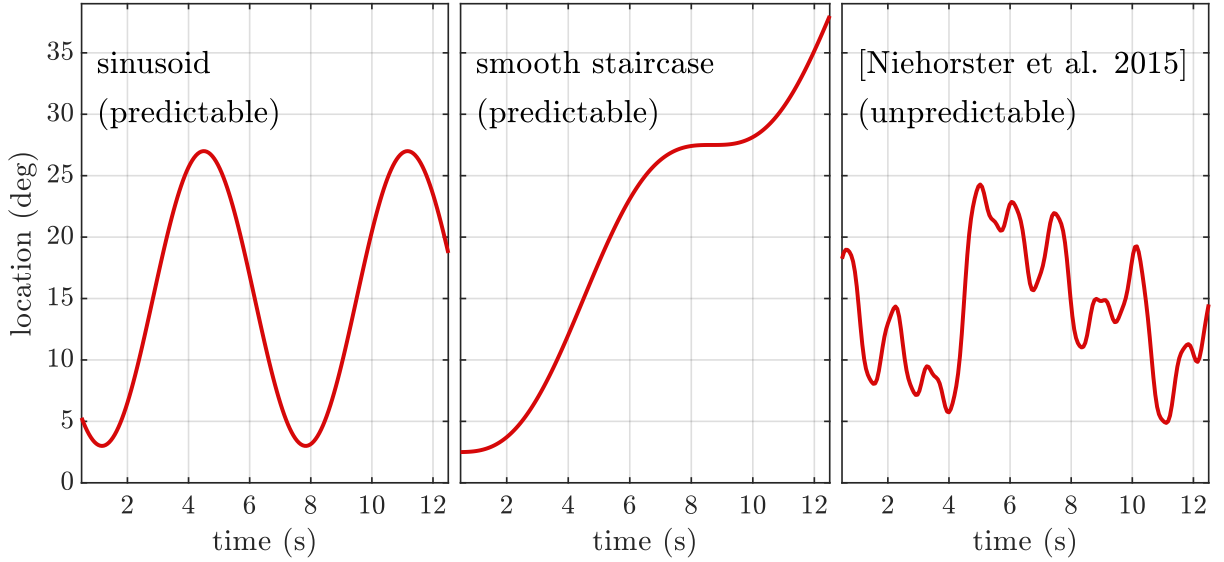


Figure 8.2: Example object motions used in the motion quality experiment. Sinusoid and smooth staircase motion (left and center) are predictable by the SPEM mechanism, while the sum of non-harmonic sinusoid (right) is unpredictable.

content, range of velocities, and type of motion. In conditions (a)–(c) the *circle* underwent predictable sinusoid motion (Figure 8.2-left) with peak velocities at  $15^\circ/\text{s}$ ,  $30^\circ/\text{s}$ , and  $45^\circ/\text{s}$ , respectively. In condition (d) the same *circle* underwent unpredictable motion (Figure 8.2-right) with mean velocity  $23^\circ/\text{s}$ . In condition (e) *ET* underwent predictable sinusoid motion (Figure 8.2-left) with peak  $15^\circ/\text{s}$ . Finally, in condition (f) *panorama* underwent a predictable motion following a soft staircase function (Figure 8.2-right):  $\theta(t) = 15^\circ (\sin(2\pi t)/2\pi t + t)$ , peak velocity at  $30^\circ/\text{s}$ . For unpredictable motion I used the same function as Niehorster et al. [2015] (the sum of non-harmonic sinusoid motions with randomised phases):

$$\theta(t) = 17^\circ \sum_{i=1}^7 a_i \sin(2\pi\omega_i t + \rho_i), \quad (8.1)$$

where  $\theta(t)$  is the horizontal object location at time  $t$ ,  $a_i = \{2, 2, 2, 2, 2, 0.2, 0.2\}$ , and  $\rho_i = \{0.1, 0.14, 0.24, 0.41, 0.74, 1.28, 2.19\}$ .

## Participants, Task

Eleven participants aged 20-42, one female ten male, with normal or corrected-to-normal vision took part. The quality was measured using a pairwise comparison protocol because of its relative simplicity and speed [Perez-Ortiz and Mantiuk 2017]. For each trial, participants were asked to select the monitor which has higher visual quality; *i.e.* the one with *sharp details and smooth motion*. Each participant performed 600 comparisons (6600 in total). During training, the researcher highlighted key differences in sharpness

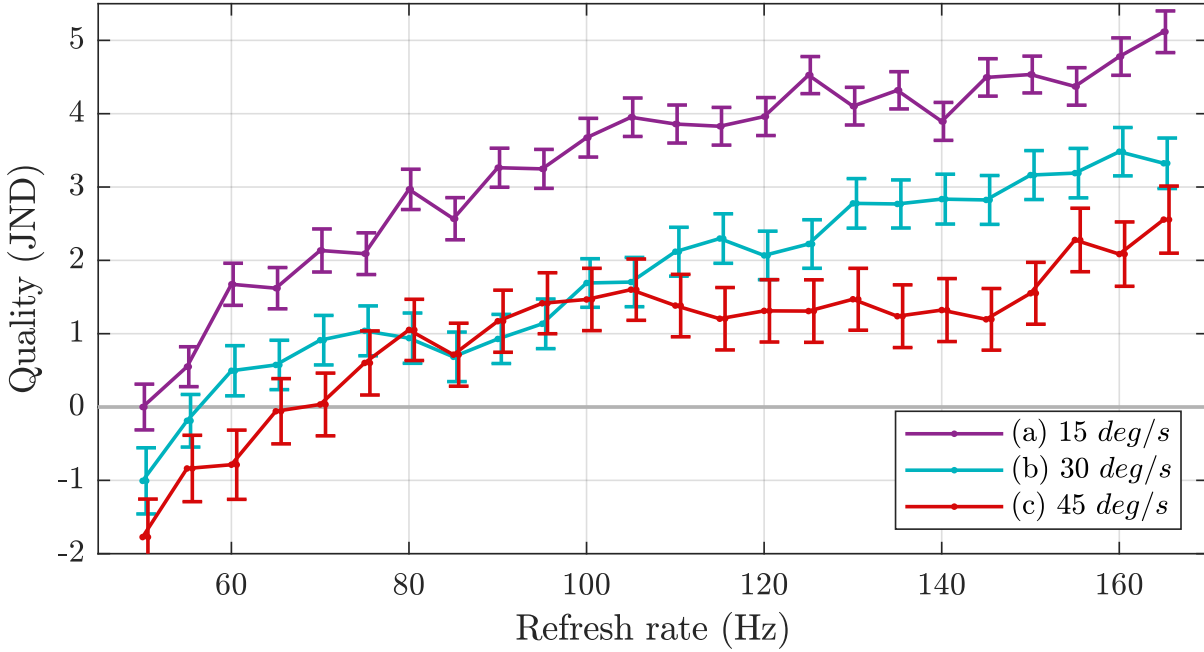


Figure 8.3: Quality across different velocities. +1 JND distance indicates preference by 75% of the population; error bars indicate 75% confidence intervals. Quality increases with refresh rate, but the increase slows down above 100 Hz. Lower velocities are perceived to be of higher quality. This is expected as higher velocities induce more blur and requires higher refresh rates to reproduce.

and motion smoothness in side-by-side 30 Hz vs. 120 Hz animations on content that was later not used in the experiment.

## Sampling

In order to efficiently utilise observers’ time and obtain the most accurate scale possible, active sampling was used [Ye and Doermann 2014; Glickman and S. 2005; Chen et al. 2013]. The next comparison was always chosen to deliver the most information, *i.e.*, the one that have would have the highest impact on the posterior distribution.

## Unified velocity scale

All 11 observers performed 420 comparisons within each condition (4620 across all participants). To establish reliable quality differences between different velocities, 180 additional comparisons were collected across velocities for conditions (a)–(c) (velocities 15 deg/s, 30 deg/s, 45 deg/s). The two measurements together enable obtaining a unified quality scale, taking into account both the refresh rate and the velocity of the object. Since JNDs are relative, the quality of the lowest measured refresh rate was set to 0 JND. To show the relative difference between the velocities in the *circle* animation, the quality of 15 deg/s at 50 Hz was set at 0 JND.

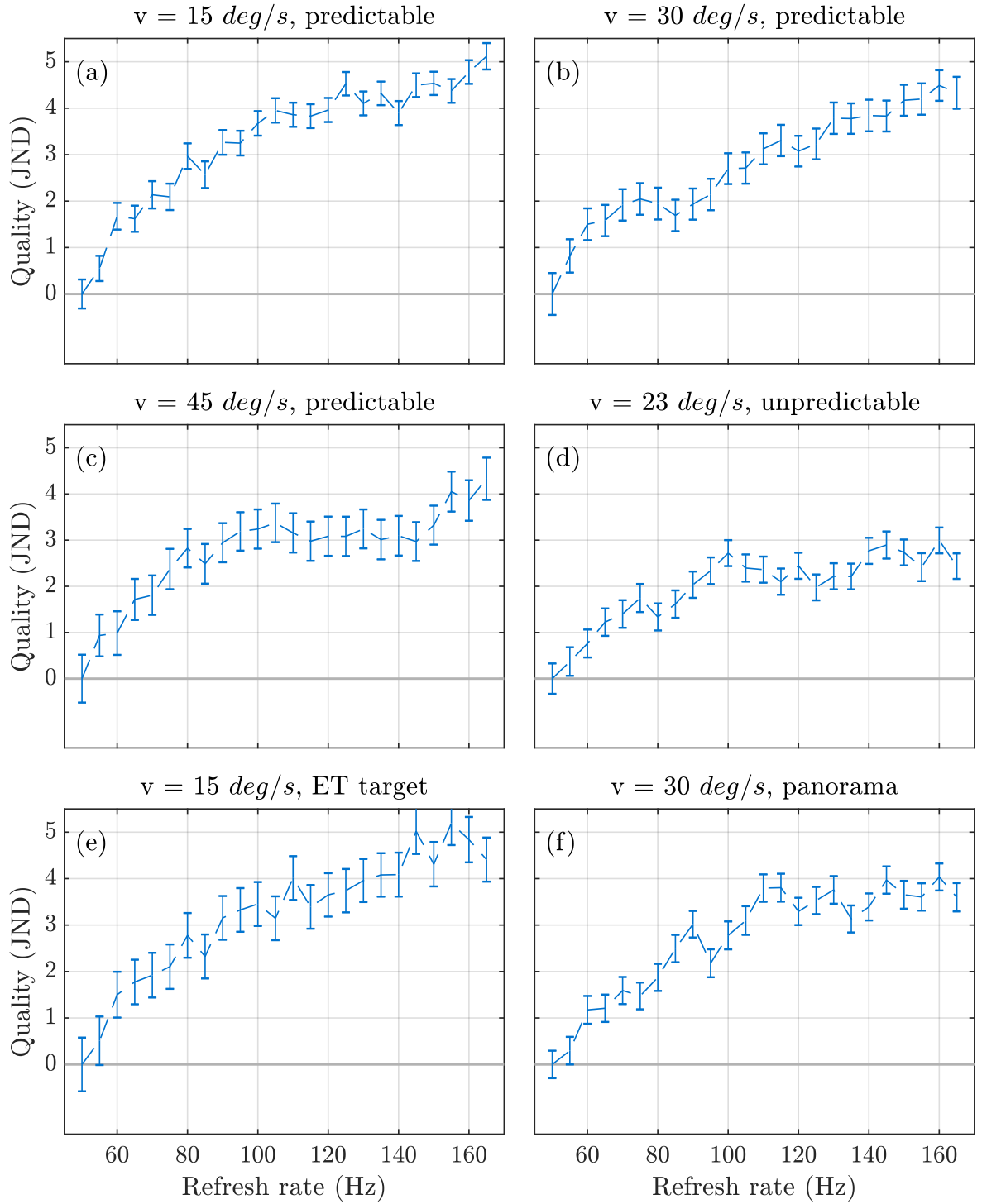


Figure 8.4: Results of the motion quality experiment for all six animations. (a-d): circle animation with different velocities; predictable and unpredictable motion. Velocity in title indicates the maximum velocity during the animation. Each curve was anchored such that 50 Hz corresponds to 0 JND.

## Results

Figure 8.3 shows the result of the cross-velocity scaling for predictable stimuli. The overall shape is consistent with most previous expectations: higher refresh rates imply less-perceivable motion artefacts, but differences above 100 Hz are increasingly more difficult to observe. In this region of refresh rates, the dominant motion artefact is blur — an artefact that is diminishing with refresh rate. There is also a clear preference for lower velocities; the explanation here is two-fold: (1) higher velocities produce more motion blur when displayed at a fixed refresh rate, and (2) content is easier to see at lower velocities, so the visual system might have an implicit preference for those. Point (2) can also explain why the velocities differ more in quality at high than at low refresh rates. At high refresh rates, when motion blur is small, the observers are picking slower motion. However, at low refresh rates, when judder is a dominant artefact, the velocities are more difficult to differentiate (all motion looks bad) and the differences in quality are becoming smaller.

For all six scenes, the predictability of motion influences the shape of the quality curve Figure 8.4. On the other hand, image content does not seem to be a strong factor, as the quality curves for circles and panorama (15 deg/s) look comparable. Similar observations can be made about circles and panorama (30 deg/s). From this, velocity and the predictability of the motion can be identified as the key factors of the model. I propose that the quality measured in this experiment can be explained by motion blur and judder.

## 8.2 A perceptual model for motion quality

In this section I present a perceptual visual model which predicts perceived quality based on the effects of refresh rate, resolution, tracked object velocity and movement type. Later, I show that the model explains experimental data, and in Section 8.4, demonstrate how it can be used to actively control the refresh rate and resolution of rendering.

Formally, I define the content-independent quality difference between rendering on display A and display B, each using different spatial resolution and refresh rates:

$$\Delta Q(\dots) = \Delta Q(f_A, R_A, f_B, R_B, v, \tau), \quad (8.2)$$

where the quality difference  $\Delta Q$  is a function of display refresh rate  $f$  (Hz), image resolution  $R$  (pixels per degree; ppd), velocity of motion  $v$  (deg/s), and predictability of motion  $\tau$  (binary input, *predictable* or *unpredictable* by SPEM). The unit of the  $Q$  function is JND. When a display is rendering at a reduced resolution, I assume an image is up-sampled to the full screen size using a bi-linear filter.

For an overview of the proposed model pipeline, see Figure 8.5. I approximate  $\Delta Q$  as

the weighted sum of three components:

$$\begin{aligned}\Delta Q(\dots) = & w_P \Delta Q_P(f_A, R_A, f_B, R_B, v, \tau) + \\ & w_O \Delta Q_O(R_A, R_B) + \\ & w_J \Delta Q_J(f_A, f_B, v, \tau); \end{aligned} \quad (8.3)$$

*i.e.* , the amount of blur in the direction of, or parallel to the motion ( $\Delta Q_P$ ), blur orthogonal to the motion, determined by the spatial resolution of the display ( $\Delta Q_O$ ), and the judder or non-smoothness of the motion ( $\Delta Q_J$ ). The following sections describe the three steps to derive  $\Delta Q_O$  and  $\Delta Q_P$ , then Section 8.2.4 describes the model for  $\Delta Q_J$ .

### 8.2.1 Blur due to spatio-temporal resolution and eye motion

The first step is to determine the loss of quality caused by the motion blur and the reduction of resolution. I separate the effect of refresh rate and resolution into three blur components: display hold-type blur ( $b_D$ ), eye motion blur ( $b_E$ ), and spatial blur due to the finite screen resolution ( $b_R$ ). I express the amount of each blur as the width of either a box or a triangle filter in visual degrees.

#### Hold-type blur ( $b_D$ )

When the eye follows a moving object, its motion is continuous, whereas LCD displays can only present a sequence of discrete samples (frames) at a finite refresh rate. As discussed in Chapter 4, current LCD displays do not necessarily emit a constant amount of light throughout a frame. However, as the transition periods have been decreasing in the recent years, and the exact transition profiles are complex, I follow Klompenhouwer et al. [2004], and approximate hold-type blur with a box filter of the width (in visual degrees):

$$b_D = \frac{v}{f}, \quad (8.4)$$

where  $v$  is the object velocity in degrees per second and  $f$  is the refresh rate in Hz.

#### Eye motion blur ( $b_E$ )

When the eye follows an object with SPEM, the tracking is imperfect. As discussed in Section 2.6.3.3, the difference between the object velocity and the gaze velocity is proportionate to the object velocity [Daly 1998]. Hence such blur can be also modelled as a box filter with width:

$$b_E = p_a v + p_b, \quad (8.5)$$

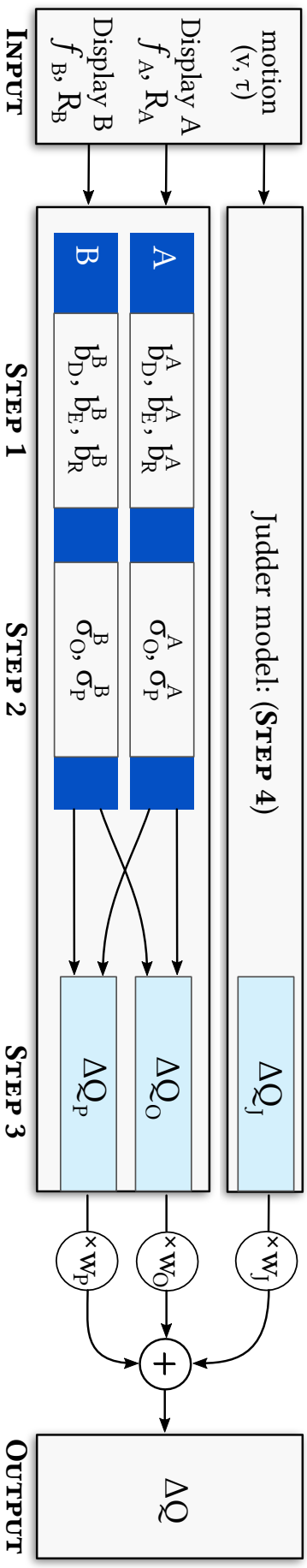


Figure 8.5: Schematic diagram of the model predicting the quality difference between two refresh rates ( $f_A, f_B$ ) and resolutions ( $R_A, R_B$ ) assuming the same motion for both refresh rates. Step 1: separately for  $A$  and  $B$ , compute blur factors due to hold-type display, eye motion, and resolution. Step 2: transform to blur kernels orthogonal and parallel to the motion. Step 3: Compute difference between  $A$  and  $B$ , and apply non-linearity (CSF, psychometric function, probability to JND transform) to find  $\Delta Q_O$  and  $\Delta Q_J$ . Step 4: Quality differences due to judder. Each step is described in the corresponding subsections of Section 8.2.

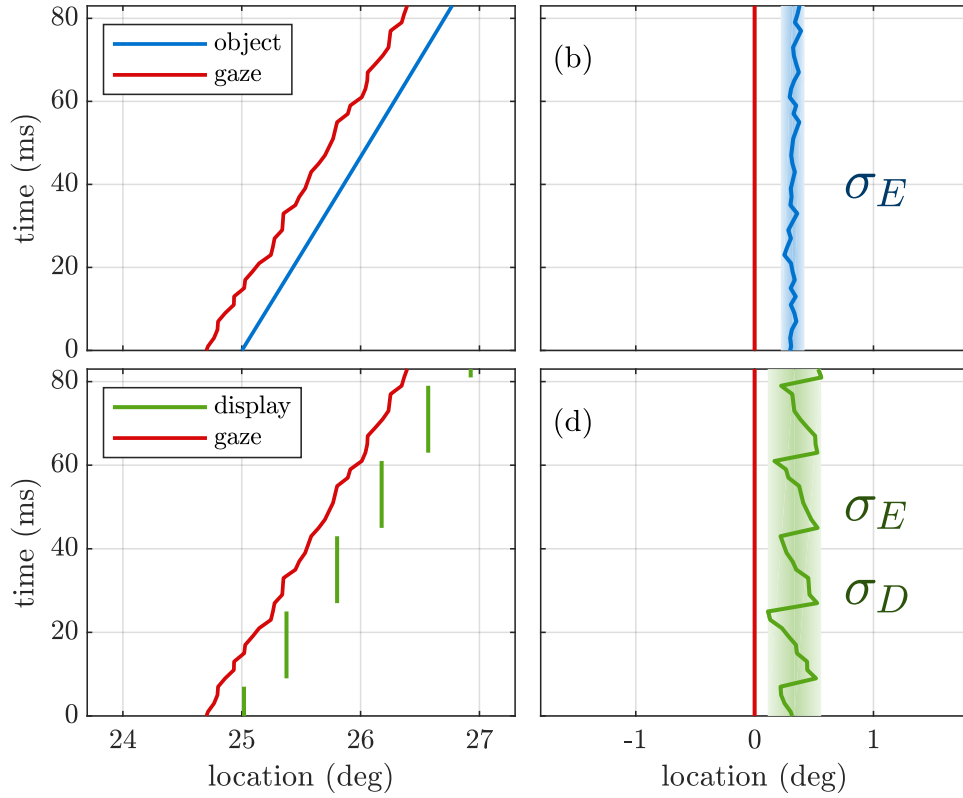


Figure 8.6: Combining eye blur ( $b_E$ ) and hold-type display blur ( $b_D$ ). Left column shows object and gaze location in absolute screen co-ordinates; right column shows the same data relative to gaze (i.e. retinal location). Object location (top row) is followed by imperfect eye motion. This introduces eye motion blur, the magnitude of which can be estimated with  $\sigma_E$ . Displayed object location is also affected by display hold-type behaviour ( $b_D$ ). The combined effect of these are shown in the bottom row. Data based on eye tracker measurements (Section 8.3.1) on 55 Hz monitor.

where  $p_a$  and  $p_b$  are constant coefficients. I assume this eye motion blur to be independent of the display refresh rate, but expect it to vary with the predictability of motion. Therefore  $p_a$  and  $p_b$  are different for predictable and unpredictable motions, as demonstrated with experimental data in Section 8.3.1.

### Spatial resolution blur ( $b_R$ )

With the general use of bilinear filters for up-sampling images in real-time graphics, the blur due to reduced spatial resolution is well-modelled by a triangle filter with average width  $b_R$  (or base width  $2b_R$ ). Given the angular resolution  $R$  in pixels per visual degree, the width of the filter is

$$b_R = \frac{1}{R}. \quad (8.6)$$

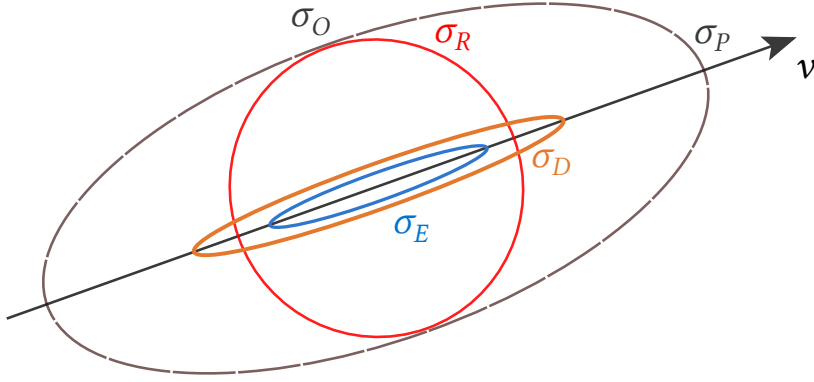


Figure 8.7: Blur is anisotropic. For a given motion in the direction  $v$ , I propose distinguishing between motion-parallel blur ( $\sigma_P$ ) and motion-orthogonal blur ( $\sigma_O$ ). Motion-parallel blur ( $\sigma_P$ ) consists of resolution reduction blur ( $\sigma_R$ ; red), eye motion blur ( $\sigma_E$ ; blue), and hold-type display blur ( $\sigma_D$ ; orange). Motion-orthogonal ( $\sigma_O$ ) blur consists only of blur due to resolution reduction ( $\sigma_R$ ; red).  $\sigma_D$  and  $\sigma_E$  ellipses are drawn for visualisation only, as I assume these sources of blur to be one-dimensional.

### 8.2.2 Motion-parallel and orthogonal blur

To simplify the combination of different blur types, I approximate each blur component with a Gaussian filter (see Figure 8.6). A box filter can be approximated with a Gaussian filter of the standard deviation:

$$\sigma = \frac{w}{\pi}, \quad (8.7)$$

where  $w$  is the width of the box filter. This implies:

$$\sigma_D = \frac{v}{\pi f}, \quad \sigma_E = \frac{p_a v + p_b}{\pi}. \quad (8.8)$$

The triangle filter, used to model the resolution reduction, can be considered as the convolution of two box filters with base width  $b_R$ . The standard deviation of this combined kernel is then

$$\sigma_R = \sqrt{\left(\frac{b_R}{\pi}\right)^2 + \left(\frac{b_R}{\pi}\right)^2} = \frac{\sqrt{2}b_R}{\pi}. \quad (8.9)$$

Eye-motion blur ( $b_E$ ), and hold-type blur ( $b_D$ ) will blur the image only in the direction of motion, but lowering spatial resolution ( $b_R$ ) will blur the image equally in all directions. Because of that, I separately compute the blur that is parallel (P) to the direction of motion and the one that is orthogonal (O) to the direction of motion, as shown in Figure 8.7.

The blur in the direction parallel to motion ( $\sigma_P$ ) is given by the convolution of individual components:

$$\sigma_P = \sqrt{\sigma_E^2 + \sigma_D^2 + \sigma_R^2}; \quad (8.10)$$

and the blur that is orthogonal to the direction of motion ( $\sigma_O$ ) is affected only by the



resolution reduction:

$$\sigma_O = \sigma_R. \quad (8.11)$$

### 8.2.3 From $\sigma$ to quality

Blur introduced by eye motion, hold-type blur, and spatial resolution will result in the loss of sharpness. To quantify this in terms of loss of perceived quality, the physical amount of blur is mapped to the perceived quality difference in JND units. The proposed blur quality function is inspired by the energy models of blur detection [Watson and Ahumada 2011]. Such mapping is applied to the orthogonal ( $\sigma_O$ ) and parallel ( $\sigma_P$ ) components of the anisotropic blur separately, resulting in two independent quality values ( $Q_O$  and  $Q_P$ ).

As we are interested in content-independent predictions, the model assume the worst-case scenario: a pixel-wide line, which is the discrete approximation of an infinitesimally thin line (Dirac delta function  $\delta(x)$ ). At the limit this contains uniform energy across all spatial frequencies. When convolved with a Gaussian blur kernel  $\sigma$  in the spatial domain, the resulting image is a Gaussian function with standard deviation  $\sigma$ .

The Fourier transform of this signal is also a Gaussian, given by:

$$m(\omega; \sigma) = \exp(-2\pi^2\omega^2\sigma^2) \quad (8.12)$$

where  $\omega$  is in cpd. To account for the spatial contrast sensitivity of visual system, the Fourier coefficients are modulated with the CSF

$$\tilde{m}(\omega, \sigma) = \text{CSF}(\omega) m(\omega; \sigma), \quad (8.13)$$

where CSF is Barten's CSF model with the recommended standard observer parameters and the background luminance of 100 cd/m<sup>2</sup> [Barten 2004].

To compute the overall energy in a distorted signal, a range of frequencies are sampled an octave apart ( $\omega_i = \{1, 2, \dots, 64\}$  [cpd]), The blur energy is then

$$E_b(\sigma) = \sum_i \left( \frac{\tilde{m}(\omega_i; \sigma)}{\tilde{m}_{t,b}} \right)^{\beta_b}. \quad (8.14)$$

where  $\tilde{m}_{t,b}$  is the threshold parameter and  $\beta_b$  is the power parameter of the model. Both of these are fitted to psychophysical data in Section 8.3.3.

Energy differences can be interpreted as quality differences, yielding:

$$\begin{aligned} \Delta Q_P &= E_b(\sigma_P^A) - E_b(\sigma_P^B), \\ \Delta Q_O &= E_b(\sigma_O^A) - E_b(\sigma_O^B), \end{aligned} \quad (8.15)$$

substituting in the standard deviations of the blur components for  $A$  and  $B$ , in the

directions parallel (P) and orthogonal (O) to SPEM.

To gain further intuition as to why an energy model is suitable to predict JND differences, let us consider the probability of selecting condition  $A$  over condition  $B$  ( $\mathbb{P}(A \succ B)$ ). This preference probability is commonly assumed to follow the psychometric function as a function of signal contrast or energy. However, the probabilities obtained by this discrimination model do not provide an intuitive, uniform quality scale. Hence, probabilities are converted to JND units under Thurstone Case V:

$$Q(A, B) = \sigma \Phi^{-1}(\mathbb{P}(A \succ B)), \quad (8.16)$$

where  $\Phi^{-1}$  is the inverse cumulative standard normal distribution. The choice of  $\sigma$  determines the relationship between distances in the quality scale and probabilities. Setting  $\sigma = 1.4826$  ensures that  $Q = 1$  when exactly 75% of observers select A over B. Note, however, that the cumulative distribution function itself is a psychometric function. Hence, the applied psychometric function which transforms energy to probability values is undone by the step of transforming probabilities to quality. The proposed model skips these steps and assumes that energy is a good indicator of quality under the threshold ( $\tilde{m}_{t,b}$ ) scaling.

#### 8.2.4 Judder ( $Q_J$ )

On lower refresh rates, finite sampling results in non-smooth, juddery motion. As described in Section 2.6.3.4, the visibility of judder can be predicted by transforming the signal to the frequency domain, and examining aliasing copies of the original signal (see Figure 2.7).

The location of the first aliasing copy, as shown in Figure 8.8-Left can be determined as follows: the temporal frequency (vertical axis) is equal to the refresh rate; the spatial frequency ( $\rho$ , horizontal axis) is:

$$\rho = \frac{f}{v}. \quad (8.17)$$

Given two refresh rates  $f_A$  and  $f_B$ , the same energy model architecture is employed as for blur. The unit signal is modulated with the spatio-temporal contrast sensitivity of the eye (stCSF), and normalized by a threshold modulation:

$$E_j(f, v) = \left( \frac{\text{stCSF}(\rho, f)}{\tilde{m}_{t,j}} \right)^{\beta_J}, \quad (8.18)$$

where  $\beta_J$  is the power parameter for judder, and  $\tilde{m}_{t,j}$  is the threshold for judder. The threshold is fitted separately for predictable and unpredictable motion. stCSF is Kelly's spatio-temporal CSF [Kelly 1979]; however, to account for the finite width of the alias, I

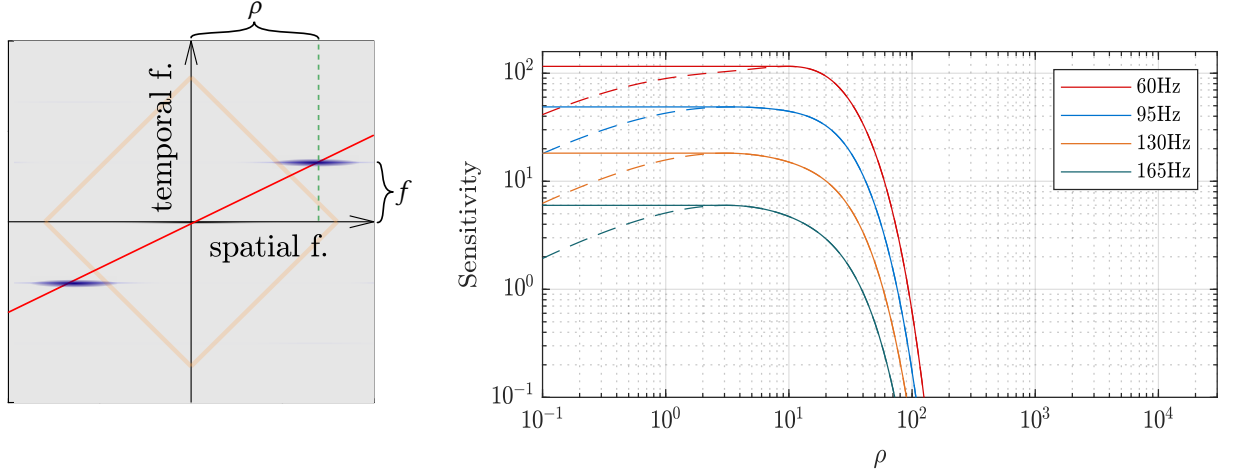


Figure 8.8: Left: the visibility of judder artefacts are determined by the location of the first aliasing copy in the frequency domain (highlighted in blue). The peak of the aliasing copy lies on the red line which in turn is determined by the spatial frequency and the object velocity. Right: spatio-temporal CSF used in the judder model. Kelly’s model predicts lower sensitivity values at low spatial frequencies (dashed); in the proposed model, I clamp this conservatively (solid). Colors show different temporal frequencies.

use a truncated low-pass stCSF, as shown in Figure 8.8-Right. Similarly as for the blur, I express the quality difference due to judder as the difference of energy:

$$\Delta Q_J = E_j(f_A, v) - E_j(f_B, v). \quad (8.19)$$

## 8.3 Model calibration

To determine the free parameters of the proposed model, I collected further data on eye motion ((Experiment 8.2) and perceived judder ((Experiment 8.3).

### 8.3.1 Retinal blur due to motion (Experiment 8.2)

Eye motion blur is caused by the differences between object and gaze motion. Daly et al. [1998] suggested that the difference in velocity and therefore also blur amount ( $b_E$ ) can be modelled as a linear function of object velocity within the SPEM-tracking range. There is, however, little data on how this function might change with unpredictable eye motion, and how to incorporate interaction with display refresh rates. To explore this problem and to fit the linear parameters ( $p_a, p_b$ ) of the proposed model described in Section 8.2.1, I measured the eye’s ability to follow predictable and unpredictable objects with an eye-tracker.

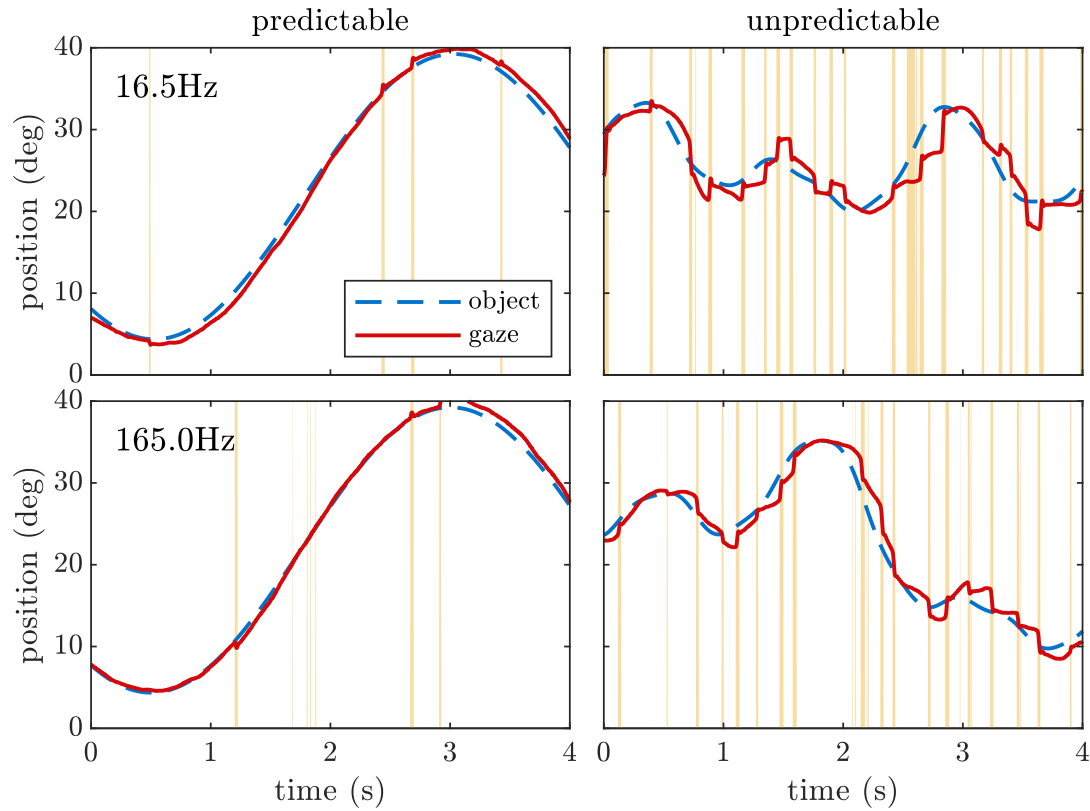


Figure 8.9: Traces of gaze location during SPEM of predictable (left) and unpredictable (right) objects at different refresh rates. Vertical yellow lines show interruptions in SPEM (saccades). Unpredictable motion visibly requires more correction saccades, with the gaze lagging behind object motion. Oscillations comparable to the respective display refresh rates are not visible.

( Hz)	No. saccades	pos err.(°)	pos var.(°)	gain	delay (s)
16.5	21.8 $\pm$ 6.8	0.70	0.50	0.60 $\pm$ 0.10	-0.01 $\pm$ 0.02
27.5	19.9 $\pm$ 4.0	0.58	0.40	0.63 $\pm$ 0.12	0.01 $\pm$ 0.01
55.0	30.6 $\pm$ 2.1	0.62	0.45	0.57 $\pm$ 0.10	0.01 $\pm$ 0.01
82.5	24.0 $\pm$ 4.2	0.70	0.58	0.59 $\pm$ 0.10	0.00 $\pm$ 0.02
165.0	20.6 $\pm$ 6.7	0.65	0.48	0.64 $\pm$ 0.11	0.01 $\pm$ 0.01
16.5	80.1 $\pm$ 12.3	1.73	1.24	0.23 $\pm$ 0.05	0.11 $\pm$ 0.02
27.5	73.6 $\pm$ 13.1	1.43	1.05	0.27 $\pm$ 0.05	0.09 $\pm$ 0.01
55.0	69.4 $\pm$ 10.0	1.33	1.01	0.30 $\pm$ 0.05	0.09 $\pm$ 0.01
82.5	72.5 $\pm$ 12.5	1.33	0.98	0.31 $\pm$ 0.06	0.08 $\pm$ 0.01
165.0	70.8 $\pm$ 12.5	1.34	1.04	0.31 $\pm$ 0.06	0.09 $\pm$ 0.00

Table 8.1: Quality of SPEM tracking. Aggregated eye tracking data for predictable (blue, top 5 rows) and unpredictable (green, bottom 5 rows) with average object velocity of 20 deg/s. Metrics described in text.

## Stimuli

The eye tracker target from Experiment 8.1 was used (bull’s eye and a cross hair; Figure 8.1). This object moved left-to-right with predictable or unpredictable motion. For predictable motion the horizontal displacement followed a sinusoidal function with the amplitude of 17° and four different frequencies to give a peak velocity of {12, 18, 24, 36} deg/s. For unpredictable motion I used the same motion as Experiment 1 (Equation 8.1). The stimuli were rendered at a range of refresh rates  $T_i = \{16.5, 27.5, 55, 60, 82.5, 120, 165\}$  Hz.

## Setup

The fixation target was displayed on the ASUS monitor with an Eyelink II eye-tracker sampling the gaze location at 500 Hz (pupil-only mode).

## Procedure

Participants were asked to follow the fixation target with their gaze with. Their head was stabilized on a chinrest 80 cm away from the monitor (field of view of 41°). Each session consisted of 30 trials, each trial lasting 20 s. A binocular 9-point calibration was performed before each trial, selecting the eye that performed better during the 9-point validation. The order of trials was randomized.

## Participants

Five participants aged 20-27 volunteered to take part in the experiments. Four participants had normal vision, while one participant wore prescription contact lenses.

## Results

Figure 8.9 shows examples of measured traces on different refresh rates for predictable and unpredictable motion. The eyetracker reported blinks, and I ignored these blink periods in the analysis. Velocities were obtained by two-point digital differentiation. Saccades were then filtered out using a threshold method when either eye velocity exceeded  $40 \text{ deg/s}$  or acceleration exceeded  $9000 \text{ deg/s}^2$ . I verified that results of the threshold method corresponded to manual labelling.

I analysed the recorded traces with metrics from [Suh et al. 2006], excluding the first five seconds of each trial and time periods of blinks. Table 8.1 shows the qualitative results of SPEM tracking averaged across trials: (1) number of saccades (2) eye position error defined as the average difference between target and gaze location measured in visual degrees, (3) eye position variability defined as the standard deviation of target-gaze difference measured in visual degrees, (4) eye gain defined as the ratio of target and gaze velocity, and (5) delay between gaze and target object, identified as the delay that gives the highest cross-correlation score of the target and gaze velocities. Note, that this definition of velocity gain differs from that of Daly et al. [Daly 1998], and as such, it is not comparable to the frequently-quoted gain value of 0.82.

The results can be summarised as follows: the eye motion is not affected by refresh rates above 27.5 Hz; but it is affected by the nature of motion (predictable vs. unpredictable). Specifically, SPEM contains significantly more saccades when tracking unpredictable motion than for predictable motion (4.89 vs. 1.56 saccades per second). The delay when tracking unpredictable motion is also significantly higher (0.092 s vs. 0.004 s). I therefore fit the same model parameters ( $p_a$ ,  $b_p$ ) for all refresh rate, but separately for the two motion types.

## Analysis

To estimate the amount of blur due to eye motion,  $b_E$ , the recorded gaze location traces are split into segments corresponding in duration to the integration time of the eye. In this analysis, I use 25 ms windows, *i.e.*, the inverse of the approximate foveal flicker fusion frequency [Simonson and Brozek 2017]. Within each integration window, I estimate eye motion blur ( $b_e$ ) as the difference between two extreme retinal positions of an object within the window, effectively measuring the width of the box filter (see Figure 8.10). To reduce measurement noise, the blur width was averaged for all windows with matching refresh rate and (binned) target velocity.

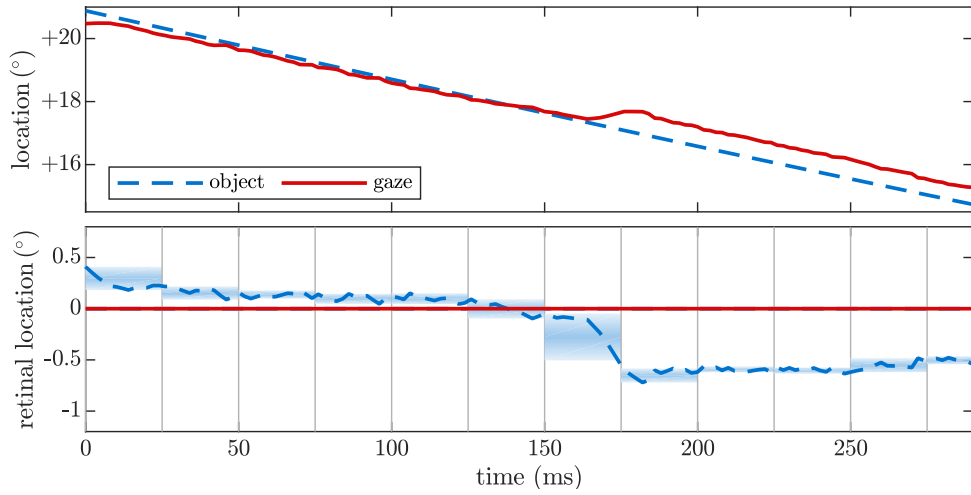


Figure 8.10: Eye blur estimation from eye tracking data. Top: gaze does not perfectly follow the target object. Bottom: retinal location is computed as the difference between gaze position and object position. The data is then split into 25-ms windows; for each window,  $b_E$  is estimated as the difference between the maximal and minimal retinal location.

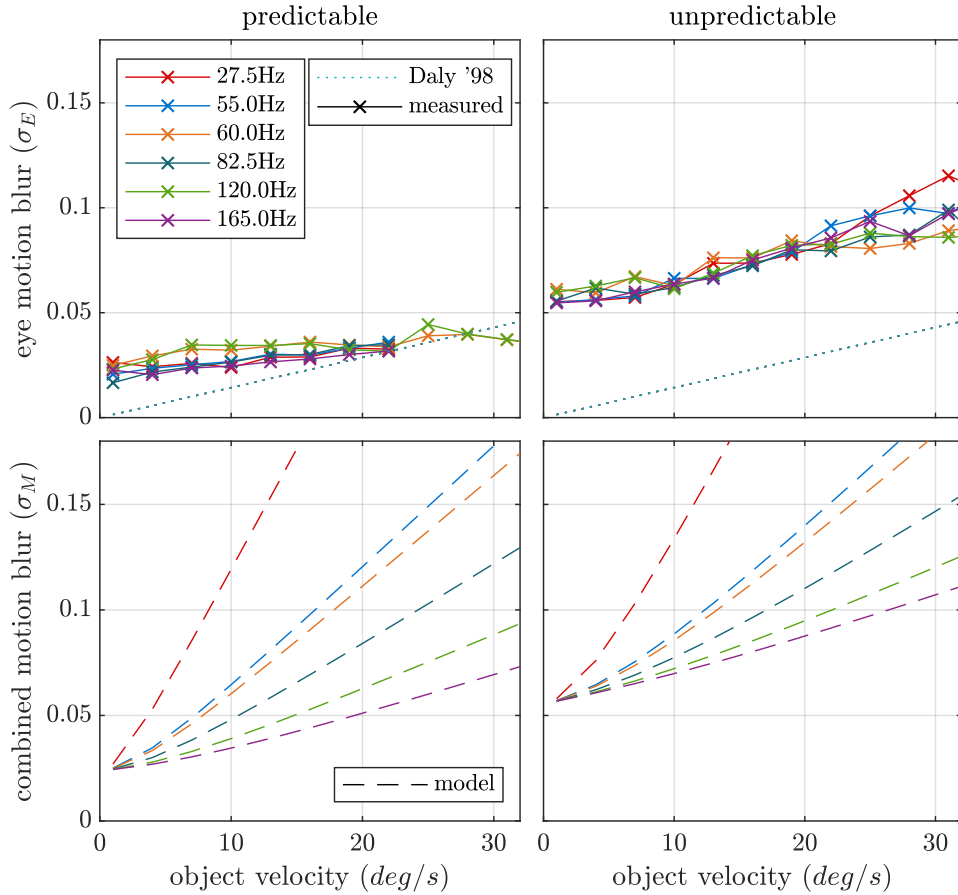


Figure 8.11: Top: blur  $\sigma$  based on eye tracker data for a range of refresh rates (different colours) and object velocities (x-axis). Blur was computed over 25ms intervals as the distance travelled by the tracked object on the retina. Dotted line: Daly's model (dotted). Bottom row: model predictions for blur taking into account both eye motion and display hold-type behaviour.

	$p_a$	$p_b$
Predictable	0.001648	0.079818
Unpredictable	0.004978	0.176820

Table 8.2: Blur model parameters. For details, see text and Figure 8.11.

### Fitting parameters $p_a$ and $p_b$

Parameters were fitted to minimise the root-mean-squared-error between model predictions from Equation 8.5 and the average blur values measured in this experiment. Figure 8.11-top indicates the measured linear relationship between object velocity and  $b_E$ . The common velocity gain of 0.82 [Daly 1998] would yield a linear gradient of  $p_a = 0.0045$  under the 25ms integration window, which the collected data for unpredictable SPEM agrees with. However, for predictable motion, results indicate much more accurate tracking (and hence less blur). For the fitted parameter values (RMSE=0.02), see Table 8.2.

### 8.3.2 Judder parameters (Experiment 8.3)

To fit the parameters of the judder model ( $\tilde{m}_{t,j}$ : energy threshold;  $\beta_j$ : power parameter), judder artefacts had to be isolated from hold-type blur and measured independently. As one cannot easily remove blur from low refresh rate and thus juddery motion, I instead opted to do the opposite: generated smooth (high refresh rate) and juddery motion (low refresh rate) and artificially introduced blur so that its amount was the same in both conditions.

#### Setup

Animations were displayed in a single ASUS display in split-screen. The distance was fixed at 80 cm using a chinrest.

#### Stimuli

Similarly to Experiment 8.1, participants observed predictable or unpredictable horizontal motion, following a fixation target or a chequered circle (Figure 8.1 right). In the split-screen setup, target 1 was rendered with refresh rate  $f$  Hz, while target 2 was rendered with  $2f$  Hz, with motion blur simulated to match the hold-type blur of  $f$  Hz. In practice this was achieved by rendering all content at  $2f$  Hz, repeating the frame in the temporal domain for target 1, and overlaying two offset frames for target 2. The spatial offset was computed as  $v/(2f)$ , where  $v$  was the actual velocity of the object.



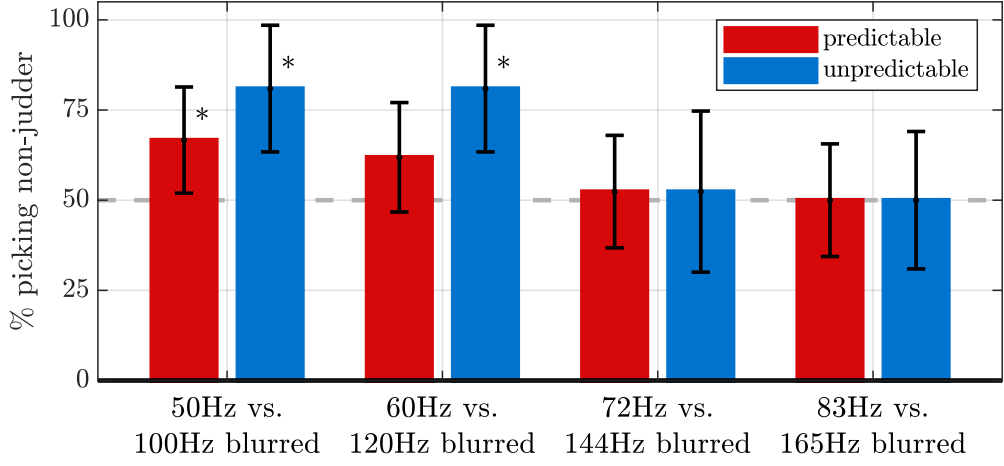


Figure 8.12: The probability of selecting the animation with reduced judder (double the refresh rate) but the same amount of motion blur, at a range of refresh rates for predictable (red) and unpredictable motion (blue). Observers were unable to tell the difference between juddery and non-juddery animations above 60 Hz. Error bars denote 95% confidence intervals.

	$\tilde{m}_{t,j}$	$\beta_J$
Predictable	218.712	2.5747
Unpredictable	165.779	

Table 8.3: Judder model parameters from Section 8.3.2.

## Task

Participants were asked to follow the fixation target with their gaze, then select the animation that provided *smoother* motion. They could view each trail for up to 20 s with the option to replay if needed. Each of the eight voluntary participants completed 108 comparisons.

## Results

The probability of detecting judder is shown in Figure 8.12. Judder was detectable for both predictable and unpredictable motion at 50 Hz and 60 Hz. At 72 Hz and 83 Hz the observers could not discriminate between the animations. This indicates that the effect of judder on quality is negligible at 72 Hz and higher refresh rates. Judder was easier to detect for unpredictable motion.

## Model fitting

As explained in Section 8.2.4, measurement results can be predicted by the energy difference in the spatio-temporal contrast sensitivity function. The best fit of Equation 8.18 to the measurement was obtained for the parameters listed in Table 8.3. The RMSE of the model predictions considering both predictable and unpredictable motion was 0.1074 JND.

$w_J$	$w_P$	$w_O$	$\tilde{m}_{t,b}$	$\beta_B$
2.218677	1.472819	1.472819	383.5854	1.83564

Table 8.4: Model parameters. For details, see text.

### 8.3.3 Fitting the quality predictions

To find the final weights of the model, I minimised the root-mean-squared error (RMSE) between the scaled output of Experiment 8.1, and the visual model for the *circle* scene. The data included 3 velocities for predictable motion and one unpredictable motion (a–d in Figure 8.13). A power of  $\beta_B = 1.83564$  and a threshold value of  $\tilde{m}_{t,b} = 383.5854$  provided the best fit with RMSE=0.312. The relative weights of judder and blur indicated that the judder component (when present) is a more significant contributor. Fitting the last parameter of the model, the weight of orthogonal blur relative to parallel blur and judder ( $w_O$ ), requires careful observation and comparison of spatial blur and motion artefacts. High  $w_O$  values bias the model to reject resolution reductions, while low  $w_O$  values result in insensitivity to orthogonal blur in slowly-moving images. The stimuli in Experiment 8.1 did not contain spatial blur orthogonal to the direction of the motion; however, the fitted value of  $w_P$  provides a reasonable starting point, as both  $\Delta Q_P$  and  $\Delta Q_O$  consider artefacts due to spatial blur. An expert observer then adjusted the relative weight of  $\Delta Q_O$  to  $\Delta Q_P$  by watching the same stimuli as in Experiment 8.1 at velocities ranging from 0 deg/s to 80 deg/s. I found that  $w_O = w_P$ , provided consistently good quality. In the next section I consider the predictions of the model, then propose and validate an application showing that the collection of parameters together can predict a good trade-off between resolution and refresh rate.

### 8.3.4 Comparison with the model of Chapiro et al.

Chapiro et al. provides a trivariate quadratic empirical model of motion quality (see Section 3.1.2). Figure 8.13 shows the predictions of their model in green. It must be noted that the maximum velocity measured in their study was 6.6 deg/s while our minimum velocity was 15 deg/s, therefore, both measurements are not directly comparable. For a better illustration, I aligned their model with the new measurements at low velocities by linear rescaling of quality predictions. Their model almost perfectly matches the new data for 15 deg/s. However, it is also clear that their model cannot extrapolate predictions for higher velocities, nor can it distinguish between predictable or unpredictable motion. For a fair comparison, I refitted their model to the new dataset by linearly rescaling the quality and reported results in Table 8.5. Their functional model does not seem to improve the fit over a fitted logarithmic function of refresh rate:  $p_1 \log(p_2 f)$ .

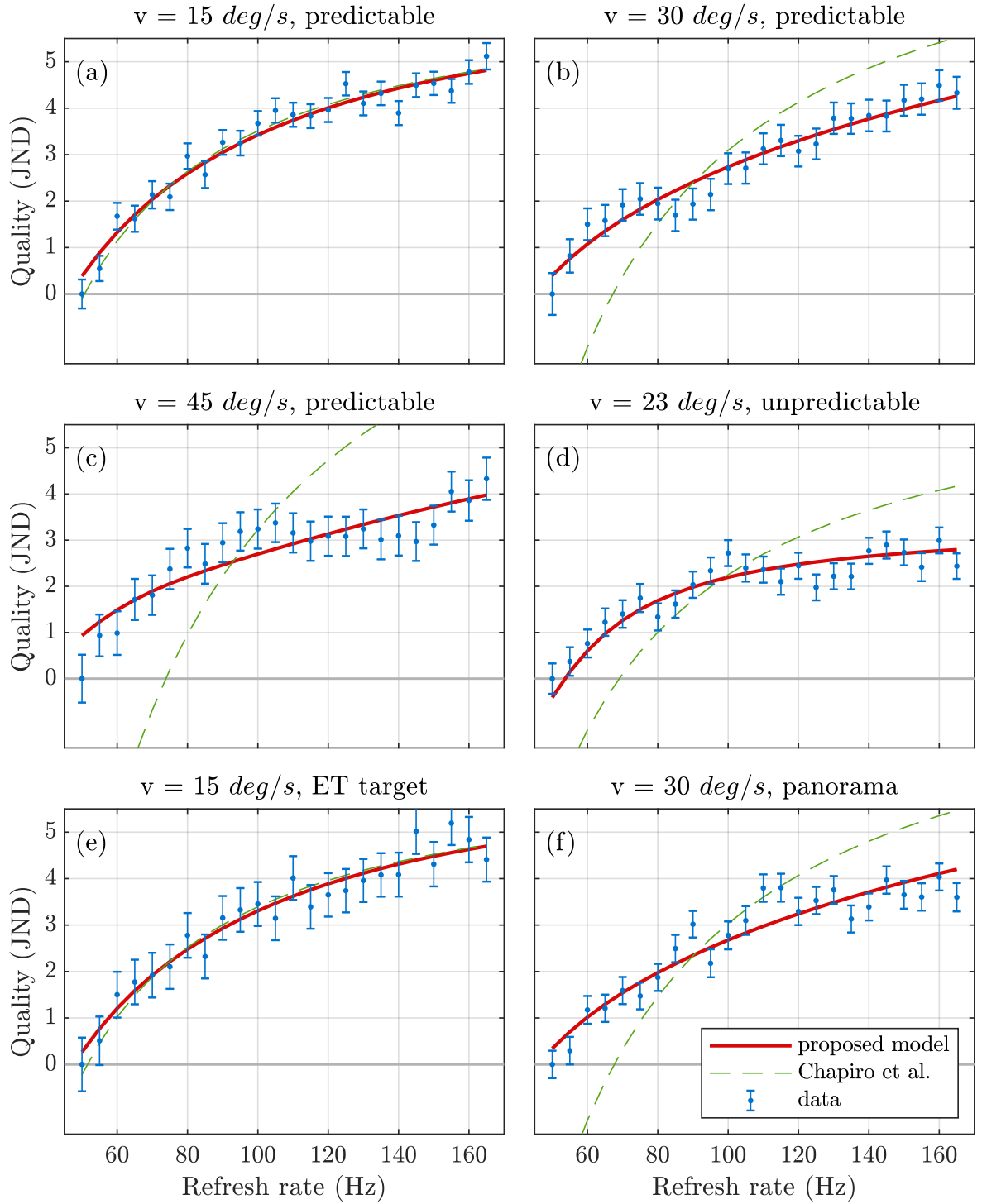


Figure 8.13: Fitted model predictions (red lines) against measurement data (blue error bars; 75% confidence). Parameters of the proposed model were fitted for the checkered circle scene (a-d). There is a distinct difference in the shape of the quality curves for different velocities (a-c) and predictable vs. unpredictable motion (d). The bottom row shows that predictions are consistent for the eye tracker target and the panorama scenes as well. The empirical model of Chapiro et al.(green dashed) provides an excellent fit for low object velocity ( $15 \text{ deg/s}$ ), but fails for higher velocities.

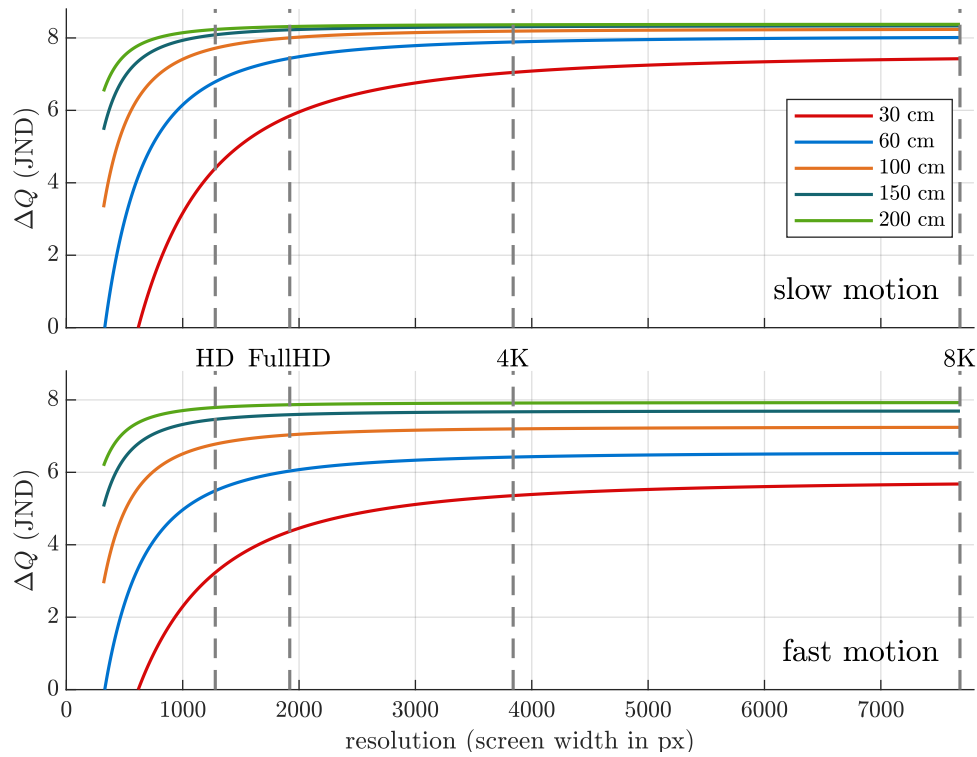


Figure 8.14: Predictions for perceived quality on a 15" display with varying resolutions (horizontal axis) and viewing distances (colours). Top: slow panning motion with the content moving horizontally across the entire screen in 6.2 s; Bottom: fast motion with content moving across the entire screen in 1.5 s. Higher resolutions bring diminishing quality gains, especially when viewed from far. Closer viewing distances also result in higher angular velocities with more visible motion artefacts.

features	train	test	(a)	(b)	(c)	(d)	(e)	(f)
Chapiro et al.	0.54	0.64	0.82	0.39	0.54	<b>0.27</b>	0.83	0.37
log	0.41	0.42	0.44	0.29	0.44	0.46	0.44	0.40
$\Delta Q_P$	0.42	0.43	0.45	0.28	0.46	0.48	0.46	0.40
$\Delta Q_P$ , pr.	0.36	0.38	0.30	0.27	0.50	0.30	0.33	0.43
$\Delta Q_J$	0.51	0.58	0.57	0.66	0.36	0.39	0.63	0.51
$\Delta Q_J$ , pr.	0.67	0.73	0.78	0.74	0.46	0.67	0.81	0.65
$\Delta Q_P$ , $\Delta Q_J$	0.36	0.35	0.28	0.36	<b>0.33</b>	0.45	0.34	<b>0.35</b>
$\Delta Q_P$ , $\Delta Q_J$ , pr.	<b>0.31</b>	<b>0.34</b>	<b>0.26</b>	<b>0.26</b>	0.42	0.29	<b>0.30</b>	0.38

Table 8.5: RMSE error (goodness of fit) for different combination of model components. Stimuli are labelled as in Figure 8.13. *pr.* indicates whether the model distinguishes between predictable and unpredictable motion. The full model provides the best fit.

### 8.3.5 Ablation study

To justify the importance of each component of the proposed visual model, I perform an ablation study. I isolate three key features: parallel quality ( $\Delta Q_P$ ), judder ( $\Delta Q_J$ ), and the isolation of unpredictable vs. predictable motion (pred). I refit the model to the circle scene for each combination of features, minimising the RMSE error in linear JND space. Goodness of fit (RMSE) is reported for the training set (the four checkered scenes), and the eye tracker target and panorama scenes as a restricted test set. Orthogonal blur cannot be separated in this study, as Experiment 8.1 did not manipulate orthogonal resolution.

The results as presented in Table 8.5 indicate that the judder model ( $\Delta Q_J$ ) on its own provides a poor fit to the quality curves (RMSE>0.51), the parallel quality factor ( $\Delta Q_P$ ) captures some trends, but cannot correctly distinguish between varying object velocities. Best predictions are provided when each model feature is enabled (RMSE=0.31). Quantifying the significance of each component is non-trivial; further model predictions are shown in Appendix C.

### 8.3.6 The effect of resolution

One advantage of the proposed model is that we can extrapolate predictions to different screen resolutions and viewing distances. Figure 8.14 shows how the perceived quality of slow (top) and fast (bottom) panning motion changes with the screen resolution (x-axis) and viewing distance. As expected, an increased screen resolution brings diminishing returns when viewed from far, and the motion looks worse from a close distance because of higher retinal velocity.

## 8.4 Model application

Limited performance budgets and transmission bandwidths mean that realtime rendering often has to compromise on the spatial resolution and temporal resolution (refresh rate) of the generated content. The typical solution to fit within a constrained rendering budget is to drive the display at a constant refresh rate and vary the rendering resolution or to render at constant resolution and vary refresh rate.

G-Sync capable monitors offer the freedom to refresh the monitor at arbitrary rates and without introducing tearing artefacts. However, under limited rendering budget, this may result in images that are sharp but juddery if the resolution is too high, or blurry but smooth animation if the resolution is too low. I propose a motion-adaptive resolution and refresh rate (MARRR) rendering algorithm, where the quality predictions of the visual model are used in real-time to establish the relative quality of different configurations of refresh rate and resolution for a fixed rendering budget bandwidth. This can be formulated as an optimisation problem from an anchor resolution  $R_\kappa$  and an anchor refresh rate  $f_\kappa$ :

$$\operatorname{argmax}_{R,f} \Delta Q(R, f, R_\kappa, f_\kappa, v, \tau) \quad \text{s.t.} \quad R f \Phi \Theta \leq B \wedge f \geq 50 Hz \quad (8.20)$$

where  $B$  is the rendering budget in pixels per second,  $\Phi$  and  $\Theta$  are the horizontal and vertical viewing angles of the monitor respectively. The optimal refresh rate will be dependent on the current object velocity, and hence, does not necessarily remain constant throughout the animation sequence. I found that the choice of the anchor point did not have a significant impact on predictions.

Figure 8.15 shows the model predictions for the ASUS display at fixed viewing distance (108 cm). For high budgets ( $> 443$  megapixels-per-second; MP/s), the model recommends keeping the refresh rate and the resolution constant up to a certain velocity and then to gradually increase the refresh rate at the cost of the resolution. The transition is more gradual for smaller rendering budgets and unpredictable motion, with slower increase in refresh rate. The predictions show a rather complex shape of the velocity-budget-rate surface.

### 8.4.1 Real-time implementation

To avoid solving an optimization problem (Equation ??) for each frame, the relation between the pixel budget ( $B$ ), velocity ( $v$ ) and the optimum refresh-rate/resolution ( $R, f$ ) can be precomputed as a look-up table (LUT). Two such LUTs, one for predictable and another unpredictable motion, are shown in Figure 8.15. In the experiments, I set the anchor frame rate to  $f_\kappa = 150 Hz$ , and sample velocity once per  $^\circ/\text{s}$ .

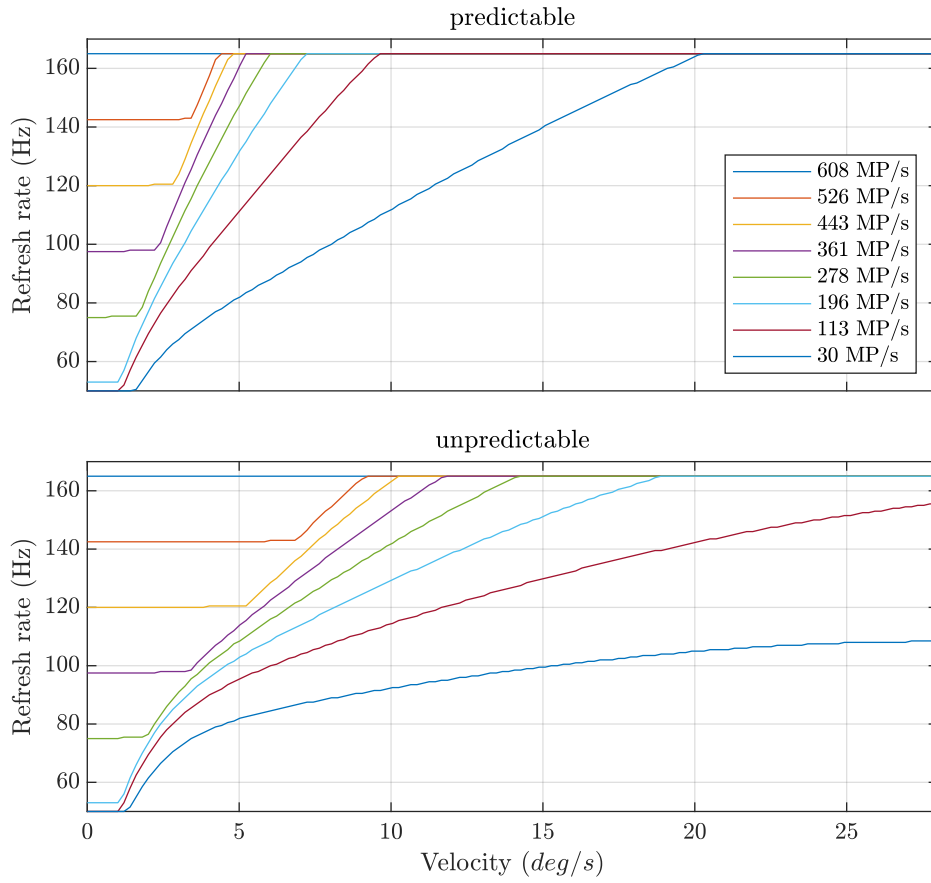


Figure 8.15: Model predictions for different rendering bandwidths (colours; measured in Megapixels per second) for predictable (top) and unpredictable motion (bottom). The plots show only the refresh rate as the resolution is determined by the fixed rendering budget.

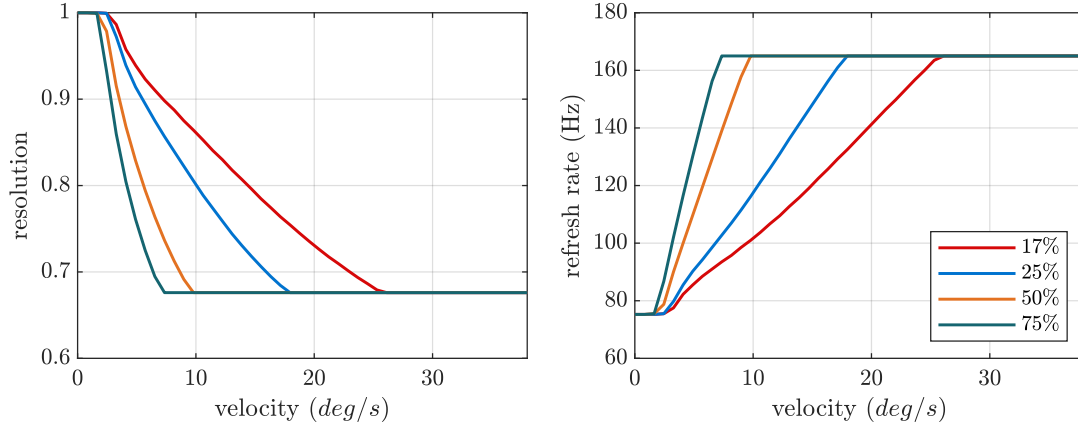


Figure 8.16: Model predictions for different display persistence values. Colours denote the percentage of the frame duration when the display is on.

### Low-persistence displays:

Although the proposed model was fitted and validated only on a high-persistence LCD display, we can extrapolate predictions to low-persistence displays, such as the ones used in VR/AR headsets. For this, it is sufficient to assume shorter integration time when estimating the amount of hold-type blur. Consequently  $(v/f)$  is replaced by  $(vp/f)$  in Equation 8.4, where  $p$  is the fraction of the frame when the display is on. The resolution/refresh rate plot in Figure 8.16 suggests that high-persistence (high percentage) demands higher refresh rates even at low velocities, whereas low persistence can keep the resolution higher under the same budget (278 MP/s). Such a model could be potentially used to dynamically control the persistence of a display to avoid visible flicker.

### Comparison with Debattista et al.:

The approach that is conceptually the closest to the one proposed here is the work of Debattista et al. [2018], discussed in Section 3.3.2. Since their model does not account for the velocity of the motion, I consider a range of potential velocities in this analysis. Assuming a fixed viewing distance of 108cm yields a field of view of  $\Phi = 30^\circ$ ,  $\Theta = 16.8^\circ$  in their setup. In Figure 8.17, I plot the optimum resolution (left) and refresh rate (right) for a given computational budget (x-axis), according to the proposed model and that of Debattista et al.. Although both models show the same trends, there are notable differences. The proposed model, intended for real-time graphics rather than cinematographic content, does not allow for refresh rates lower than 50 Hz. Its model recommends overall higher refresh-rates and lower resolutions, especially when the velocity of motion is high. However, when the budget is sufficiently large ( $>100$  MP/s) and velocities are low, the proposed model recommends higher resolutions than that of Debattista et al. This demonstrates an adaptivity to velocity.



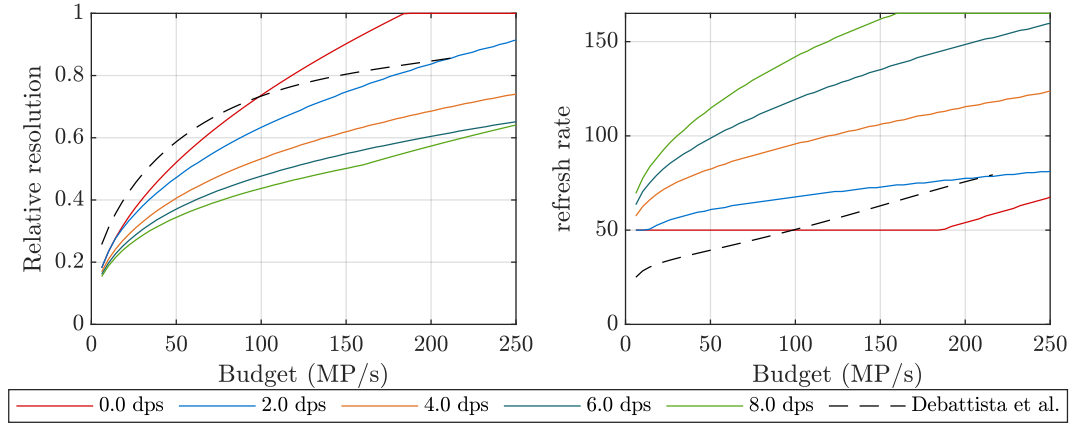


Figure 8.17: Model predictions for predictable motion for different velocities (colours) in  $\text{deg/s}$  (dps) plotted against Debattista et al. [2018] (dashed line). Assuming viewing distance of 108 cm and a field of view of  $30^\circ$ . Resolution is relative to linear image size; Refresh rate predictions are for QHD resolution.

Note that the proposed model has several further advantages, such as adapting to any viewing distance, maximal display resolution and refresh rate. In the next section I demonstrate in a psychophysical experiment that adaptive rendering based on the proposed model is preferred over Debattista et al. even within their operational range.

#### 8.4.2 Experiment 8.4: psychophysical validation

To compare MARRR with the current state-of-the-art approach, I pick three computation budgets and their corresponding refresh rates from Debattista et al. [2018], and demonstrate how MARRR produces subjectively preferred results. The selected bandwidths were  $B_i = \{28, 55, 221\} \text{ MP/s}$ . To account for potential viewing condition differences between the experiment setups, I tested three fixed refresh rates on and around the reference values reported in [Debattista et al. 2018].

##### Setup

The experiment used a 2AFC design with the same setup as in Section 1 — two G-sync capable monitors stacked on top of each other. I implemented a custom C++ OpenGL application that allowed the users to scroll across a panorama image using either a mouse or predetermined motion. On one monitor, the renderer used a single refresh rate throughout the entire animation; on the other monitor, the renderer established the optimal refresh rate (from 50 Hz to 165 Hz) frame-by-frame according to the proposed visual model. For this, a pre-computed look-up table was used with three input parameters: bandwidth, velocity, and motion predictability (Figure 8.15). The application reduced rendering resolution to meet the budget requirements. The two monitors displayed the same content



Figure 8.18: Panorama images used in the validation experiment.

but at different resolutions and refresh rates. The mouse movement was synchronised over the network.

## Stimuli

For content, four high-quality panorama images were picked (Figure 8.18). For predictable motion, the observer could pan the panoramas by moving the mouse. Such user-controlled motion is predictable and is similar to the target application, *e.g.* camera rotation in a first-person game, or camera panning in real-time strategy or simulator games.

For unpredictable motion, I used the same formula as previously (Equation 8.1). The experiment was more difficult to implement for unpredictable motion. The rapid changes in refresh rates combined with mis-predictions in the control system of G-Sync during

the unpredictable motion caused occasional skipped frames. Such motion flaw is not a limitation of the visual model or the algorithm, but the lack of ability to aid the G-Sync control system from an application side when picking the current refresh rate. To reduce these artefacts, I discretised the predicted refresh rates to integer divisors of 165 Hz: {55, 82.5, 165} Hz.

## Participants and procedure

Nine people (aged 20–40) volunteered to take part in the experiment. They all had normal or corrected-to-normal vision. Participants were asked to imagine a scenario where they were purchasing a new monitor for playing computer games, and for each trial to pick either the top or the bottom monitor based on overall visual quality (including motion and sharpness). The order of the comparisons and the presentation on the monitors were randomised. Each observer completed 81 and 27 comparisons for predictable and unpredictable motion, respectively. All participants reported only casual gaming experience (playing only a few times a month) with little-to-no exposure to high-refresh-rate monitors.

## Results

The results of the validation experiment, shown in Figure 8.19, indicate an overall preference for MARRR as compared to the fixed rates from [Debattista et al. 2018]. The difference is particularly strong for mouse-induced (predictable) motion: for all bandwidths, the proposed algorithm was picked with over a 70% probability. The trend indicates that the impact of activating adaptive rendering was lower for higher bandwidths, which is consistent with expectations. For unpredictable motion, MARRR provided better overall results, but for high-refresh-rate conditions the experiment was inconclusive. Better synchronisation capabilities with the monitor should allow for a less noisy comparison in the future.

## 8.5 Limitations

I derived this model for the assumption of the worst-case content — an pixel-wide line orthogonal to the direction of motion. This helped us to make the model content-independent so that its predictions can be pre-computed and stored as a look-up table. However, a less conservative model, considering image contents, could potentially provide better control over the resolution and refresh rate. The model also made a general assumption on the monitor brightness, which implies that different displays (especially high-luminance HDR displays) might require a different fit.

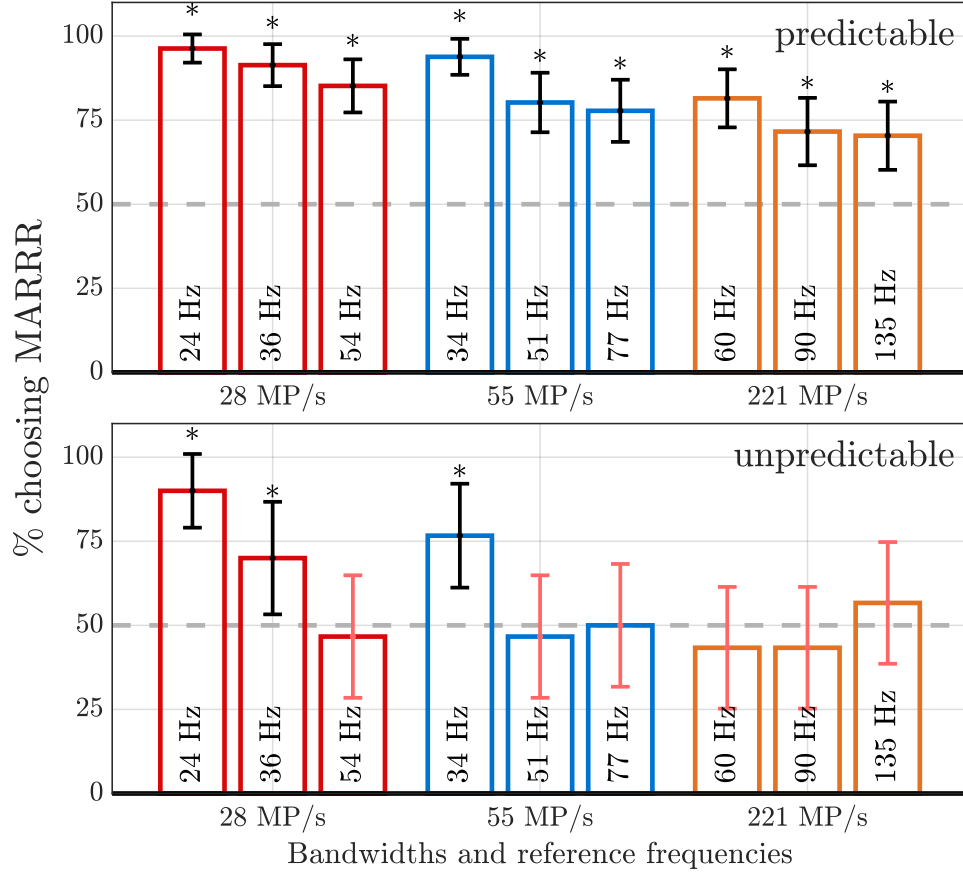


Figure 8.19: Results of validation experiment showing the percentage of participants picking the proposed adaptive MARRR algorithm over standard *constant*-resolution-and-*constant*-refresh-rate rendering, viewing predictable (top) and unpredictable (bottom) motion. Colours denote different rendering budgets. The refresh rates were selected around the predictions of [Debattista et al. 2018]. Error-bars denote 95% confidence intervals.

While the validation experiment (Section 4) showed positive results, current display technology seems to be a greatly limiting factor for validation. I anticipate that adaptive spatio-temporal resolution could have a significant impact on virtual reality, but I was unable to test this hypothesis without a VR headset supporting adaptive refresh rates. I also faced issues driving G-Sync with precision and accuracy. Operating system interrupts and other background tasks can throw the algorithm off, revealing blurry artefacts. Current monitors implement G-Sync as a control system. This means that the display is able to adapt to arbitrary refresh rates, but estimating the new refresh rate and transitioning to it takes time. A direct interface to request a specific refresh rate from G-Sync would greatly benefit the proposed technique and would allow a better evaluation especially for unpredictable motion.

The proposed algorithm requires knowledge of the velocity of the object the user is following with the gaze. In the experiments, I achieved this by instructing the user to follow a certain object, or applying the same motion vector across the whole screen. Ongoing work by my co-authors have demonstrated that MARRR can also operate in complex game setups. They achieved this by integrating the algorithm into a popular Unity game, combining user input and gaze location to derive a SPEM velocity value. More detail on this work is to follow.

Finally, while the proposed visual model generalises in principle to VR headsets, parameters were calibrated to high-persistence displays. Further psychophysical experiments would be needed to re-fit for low-persistence headset with potential extensions to account for ghosting and flicker artefacts.

## 8.6 Summary

Simple insights into the visual system are powerful. However, to create, calibrate and analyse novel graphical algorithms, we need to either rely on a new psychophysical experiment every time, or a visual metric that is known to correlate with subjective preference. In this chapter, I presented new psychophysical data on motion perception on high-frequency monitors (50–165 Hz). Then, I introduced a visual model for two prominent motion artefacts: judder and motion blur. Finally, I used this visual model to design a novel motion-adaptive resolution and refresh-rate rendering (MARRR) algorithm. The presented results also revealed that while an empirical function might offer a simpler model, a white-box approach can be still desirable, as it is more robust in novel circumstances.



---

---

# CHAPTER 9

---

## CONCLUSION

*“The curtain now rises upon the last act of our little drama, for hard-hearted publishers warn me that a single volume must of necessity have an end.”*

*Thomas Hughes*

*Tom Brown’s schooldays*

The main goal of my PhD was to demonstrate that insights into and models of temporal perception can play a crucial role in the future of computer graphics. Graphics algorithms typically run on a GPU – perhaps the most powerful part of modern computers. However, with the rapid improvement of VR and AR, the resolution, refresh rate and power demand of these headsets have shown that naïve approaches such as rendering every pixel for each frame is simply unfeasible. Researchers have reached out to the field of visual science, and have proposed algorithms with insights into visual perception. The resulting algorithms, such as foveated rendering and chroma subsampling show promising results; however, the temporal domain has received remarkably little attention. In this dissertation, I highlight the importance of the temporal domain, and suggest models and algorithms that incorporate perceptual knowledge to aid the design of future computer graphics algorithms.

### 9.1 Contributions

The main contributions of my work are as follows:

## **Temporal resolution multiplexing**

In Chapter 5, I demonstrated how a relatively simple insight into the visual system can drive the design of a novel rendering algorithm. Specifically, I exploited the eyes' limited sensitivity when it comes to high spatial and high temporal frequencies. TRM's design and integrability with existing pipelines makes it an appealing algorithm for VR. This work has received both academic recognitions (best journal paper award), and interest from industry partners who develop next-generation VR headsets.

## **Multi-scale visual models in the temporal domain**

In Chapters 6–8, I described how multi-scale visual metrics can be applied to temporal problems, such as flicker detection, and motion quality estimation. The models incorporate knowledge from the field of vision science, and fit any free parameters to results of psychophysical experiments.

## **Motion-adaptive rendering**

In the second half of Chapter 8, I presented a novel adaptive rendering algorithm which takes screen parameters (maximum resolution, viewing distance, maximum refresh rate), and on-screen motion information to account to predict the ideal trade-off between resolution and refresh rate for a fixed computational budget. The algorithm showcases how the visual models proposed earlier in the dissertation can be applied to render content efficiently.

## **Display modelling**

While perhaps not so much of an individual contribution, but certainly a crucial aspect of my work is the use of physical units to express stimuli (e.g.  $\text{cd/m}^2$  for luminance). Much of the existing literature in computer graphics and image processing relies on display-referenced pixel values, which are not robustly applicable when applied to a novel display with higher field of view, higher luminance, or just a previously uncommon viewing distance. In this dissertation, I provided accurate physical measurements of all displays involved, establishing both their luminance response and temporal characteristics. I also fitted existing display models and provided new extensions when needed.

## **9.2 Future work**

Human perception is an incredibly complex problem; the visual system has several unexpected limitations. Even within the temporal domain there are numerous directions



unexplored in this work. Since VR headsets offer a previously unseen field of view (above 110 degrees), a promising direction would be to explore temporal artefacts in the periphery. Foveated rendering is known to reduce rendering cost by up to 80%, but current metrics do not even consider the highly-prominent flicker artefacts that are just as visible in the periphery, as in the fovea. A topic of similar interest could be how motion can mask flicker artefacts. Especially in the periphery, modelling this phenomenon could help to drive the trade-off between flickering and ghosting artefacts, commonly found in modern temporal anti-aliasing (TAA) algorithms. Another interesting direction is the incorporation of colour: colour sensitivity is known to be low for high spatio-temporal signals; a limitation which could be exploited by an algorithm similar to TRM. Finally, in this dissertation, I considered only three of the four motion artefacts. Although *flicker*, *blur* and *judder* artefacts are the most prominent in modern real-time graphical applications, future studies could investigate the relative importance of *false edges*, especially in the context of low-persistence displays.

### 9.3 Final remarks

Human vision and computer graphics are inherently related: content produced by graphics algorithms is primarily “consumed” by human observers. I hence started this dissertation following the dual of the well-known computer vision argument that “vision is inverse graphics”. Following this logic, I argued that understanding the human visual system can drive the design of novel graphics algorithms. However, to come up with the perfect invertible model is simply impossible for now. We do not possess enough knowledge of the exact behaviour of the visual system – and since it is part of our brain, some argue that we shall never have a complete model. However, that should not stop graphics researchers from applying existing visual science knowledge. I believe this work has demonstrated how some understanding of the visual system can lead to crucial performance savings. At the same time, the dissertation also highlights that computer graphics research would benefit from considering the temporal dimension with more care.



---

# BIBLIOGRAPHY

- Afzal, S., Chen, J., and Ramakrishnan, K. K. (2017). Characterization of 360-degree Videos. In *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network - VR/AR Network '17*, pages 1–6, New York, New York, USA. ACM Press.
- Anthes, C., García-Hernández, R. J., Wiedemann, M., and Kranzlmüller, D. (2016). State of the art of virtual reality technology. In *IEEE Aerospace Conference Proceedings*, volume 2016-June, pages 1–19. IEEE.
- Baker, C. H. (1970). A study of the Sherrington effect. *Perception & Psychophysics*, 8(6):406–410.
- Baldwin, A. S., Schmidtman, G., et al. (2015). Modeling probability and additive summation for detection across multiple mechanisms under the assumptions of signal detection theory. *Journal of vision*, 15(5):1–1.
- Barten, P. G. J. (2004). Formula for the contrast sensitivity of the human eye. In Miyake, Y. and Rasmussen, D. R., editors, *Image Quality and System Performance*, volume 5294, pages 231–238.
- Beeler, D., Hutchins, E., and Pedriana, P. (2016). Asynchronous spacewarp. <https://developer.oculus.com/blog/asynchronous-spacewarp/>. Accessed: 2018-05-02.
- Beigbeder, T., Coughlan, R., Lusher, C., Plunkett, J., Agu, E., and Claypool, M. (2004). The effects of loss and latency on user performance in unreal tournament 2003®. In *Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games*, pages 144–151. ACM.
- Berns, R. S. (1996). Methods for characterizing CRT displays. *Displays*, 16(4 SPEC. ISS.):173–182.

- Berthouzoz, F. and Fattal, R. (2012). Resolution enhancement by vibrating displays. *ACM Transactions on Graphics*, 31(2):1–14.
- Boitard, R., Mantiuk, R. K., and Pouli, T. (2015). Evaluation of color encodings for high dynamic range pixels. In Rogowitz, B. E., Pappas, T. N., and de Ridder, H., editors, *Human Vision and Electronic Imaging XX*, volume 9394, page 93941K.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10:433–436.
- Briggs, F. (2017). Mammalian visual system organization. In *Oxford Research Encyclopedia of Neuroscience*.
- Brown, W. R. J. (1957). Color Discrimination of Twelve Observers\*. *Journal of the Optical Society of America*, 47(2):137.
- Caelli, T., Hübner, M., and Rentschler, I. (1985). The detection of phase shifts in two-dimensional images. *Perception & Psychophysics*, 37(6):536–542.
- Cao, G., Zhao, Y., Ni, R., and Kot, A. C. (2011). Unsharp masking sharpening detection via overshoot artifacts analysis. *IEEE Signal Processing Letters*, 18(10):603–606.
- Carpenter, L. (1984). The a-buffer, an antialiased hidden surface method. In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 103–108.
- Castet, E. (2009). Perception of intra-saccadic motion. In *Dynamics of visual motion processing*, pages 213–238. Springer.
- Chapiro, A., Atkins, R., and Daly, S. (2019). A luminance-aware model of judder perception. *ACM Transactions on Graphics (TOG)*, 38(5):1–10.
- Chen, H., Ha, T., Sung, J., and Han, B. (2009). 33.4: Smooth-Frame-Insertion Method for Reducing Motion Blur on OLED Panel. *SID Symposium Digest of Technical Papers*, 39(1):472.
- Chen, H., Kim, S. S., Lee, S. H., Kwon, O. J., and Sung, J. H. (2006). Nonlinearity compensated smooth frame insertion for motion-blur reduction in LCD. In *2005 IEEE 7th Workshop on Multimedia Signal Processing*, pages 1–4. IEEE.
- Chen, X., Bennett, P. N., Collins-Thompson, K., and Horvitz, E. (2013). Pairwise ranking aggregation in a crowdsourced setting. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM ’13*, pages 193–202, New York, NY, USA. ACM.

- Claypool, K. T. and Claypool, M. (2007). On frame rate and player performance in first person shooter games. *Multimedia Systems*, 13(1):3–17.
- Claypool, M. and Claypool, K. (2009). Perspectives, frame rates and resolutions. In *Proceedings of the 4th International Conference on Foundations of Digital Games - FDG '09*, page 42, New York, New York, USA. ACM Press.
- Cooper, E. A., Jiang, H., Vildavski, V., Farrell, J. E., and Norcia, A. M. (2013). Assessment of OLED displays for vision research. *Journal of Vision*, 13(12):16–16.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., and Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, 292(4):497–523.
- Daly, S., Xu, N., Crenshaw, J., and Zunjarrao, V. J. (2015). A Psychophysical Study Exploring Judder Using Fundamental Signals and Complex Imagery. *SMPTE Motion Imaging Journal*, 124(7):62–70.
- Daly, S. J. (1998). Engineering observations from spatiovelocity and spatiotemporal visual models. In Rogowitz, B. E. and Pappas, T. N., editors, *Human Vision and Electronic Imaging III*, volume 3299, pages 180 – 191. International Society for Optics and Photonics, SPIE.
- Daly, S. J. (2005). Visible differences predictor: an algorithm for the assessment of image fidelity. *Human Vision, Visual Processing, and Digital Display III*, 1666:2.
- Daly, S. J. and Feng, X. (2003). Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. In Eschbach, R. and Marcu, G. G., editors, *Color Imaging VIII: Processing, Hardcopy, and Applications*, volume 5008, page 455.
- Daly, S. J. and Feng, X. (2004). Decontouring: prevention and removal of false contour artifacts. *Human Vision and Electronic Imaging IX*, 5292:130.
- Davis, J., Hsieh, Y.-H., and Lee, H.-C. (2015). Humans perceive flicker artifacts at 500 Hz. *Scientific reports*, 5:7861.
- De Lange Dzn, H. (1952). Experiments on flicker and some calculations on an electrical analogue of the foveal systems. *Physica*, 18(11):935–950.
- de Lange Dzn, H. (1958). Research into the Dynamic Nature of the Human Fovea–Cortex Systems with Intermittent and Modulated Light II Phase Shift in Brightness and Delay in Color Perception. *Journal of the Optical Society of America*, 48(11):784.

- De Simone, F., Frossard, P., Wilkins, P., Birkbeck, N., and Kokaram, A. (2017). Geometry-driven quantization for omnidirectional image coding. In *2016 Picture Coding Symposium, PCS 2016*, pages 1–5. IEEE.
- Debattista, K., Bugeja, K., Spina, S., Bashford-Rogers, T., and Hulusic, V. (2018). Frame rate vs resolution: A subjective evaluation of spatiotemporal perceived quality under varying computational budgets. *Computer Graphics Forum*, 37(1):363–374.
- Denes, G., Ash, G., and Mantiuk, R. (2019a). A visual model for predicting chromatic banding artifacts. *Electronic Imaging*.
- Denes, G., Maruszczczyk, K., Ash, G., and Mantiuk, R. K. (2019b). Temporal Resolution Multiplexing: Exploiting the limitations of spatio-temporal vision for more efficient VR rendering. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2072–2082.
- Didyk, P., Eisemann, E., Ritschel, T., Myszkowski, K., and Seidel, H.-P. (2010a). Apparent display resolution enhancement for moving images. *ACM Transactions on Graphics*, 29(4):1.
- Didyk, P., Eisemann, E., Ritschel, T., Myszkowski, K., and Seidel, H.-P. P. (2010b). Perceptually-motivated Real-time Temporal Upsampling of 3D Content for High-refresh-rate Displays. *Computer Graphics Forum*, 29(2):713–722.
- DoVale, E. (2017). High Frame Rate Psychophysics: Experimentation to Determine a JND for Frame Rate. *SMPTE Motion Imaging Journal*, 126(9):41–47.
- Eilertsen, G., Mantiuk, R. K., and Unger, J. (2016). A high dynamic range video codec optimized by large-scale testing. In *Proceedings - International Conference on Image Processing, ICIP*, volume 2016-Augus, pages 1379–1383. IEEE.
- Elliott, C. H. B., Credelle, T. L., Han, S., Im, M. H., Higgins, M. F., and Higgins, P. (2005). Development of the PenTile Matrix™ color AMLCD subpixel architecture and rendering algorithms. *Journal of the Society for Information Display*, 11(1):89.
- Feng, X.-f. (2006). CD motion-blur analysis, perception, and reduction using synchronized backlight flashing. In *Human Vision and Electronic Imaging XI*, volume 6057, page 60570M.
- Franck Diard (2015). Adaptive resolution DGPU rendering to provide constant framerate with free IGPU scale up.
- Fredericksen, R. E. and Hess, R. F. (1998). Estimating multiple temporal mechanisms in human vision. *Vision Research*, 38(7):1023–1040.

- Fujibayashi, A. and Boon, C. S. (2008a). A masking model for motion sharpening phenomenon in video sequences. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E91-A(6):1408–1415.
- Fujibayashi, A. and Boon, C. S. (2008b). Application of motion sharpening effect in video coding. In *Proceedings - International Conference on Image Processing, ICIP*, pages 2848–2851. IEEE.
- Geisler, W. S. and Perry, J. S. (1998). Real-time foveated multiresolution system for low-bandwidth video communication. In Rogowitz, B. E. and Pappas, T. N., editors, *Human Vision and Electronic Imaging III*, volume 3299, pages 294 – 305. International Society for Optics and Photonics, SPIE.
- Georgeson, M. A. (1987). Temporal properties of spatial contrast vision. *Vision Research*, 27(5):765–780.
- Glickman, M. and S., J. (2005). Adaptive paired comparison design. *Journal of Statistical Planning and Inference*, 127(1–2):279–293.
- Guenter, B., Finch, M., Drucker, S., Tan, D., and Snyder, J. (2012). Foveated 3d graphics. *ACM Transactions on Graphics (TOG)*, 31(6):164.
- Hainich, R. and Bimber, O. (2016). *Displays*. CRC Press, Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
- Hammett, S. T. and Bex, P. J. (1996). Motion sharpening: Evidence for the addition of high spatial frequencies to the effective neural image. *Vision Research*, 36(17):2729–2733.
- Hammett, S. T., Georgeson, M. A., and Gorea, A. (1998). Motion blur and motion sharpening: Temporal smear and local contrast non-linearity. *Vision Research*, 38(14):2099–2108.
- Hartmann, E., Lachenmayr, B., and Brettel, H. (1979). The peripheral critical flicker frequency. *Vision Research*, 19(9):1019–1023.
- Hoffman, D. M., Karasev, V. I., and 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):255.
- Holzman, P. S., Levy, D. L., and Proctor, L. R. (1976). Smooth Pursuit Eye Movements, Attention, and Schizophrenia. *Archives of General Psychiatry*, 33(12):1415–1420.

- Ishan Goradia, Jheel Doshi, and Lakshmi Kurup (2014). A Review Paper on Oculus Rift & Project Morpheus. *International Journal of Current Engineering and Technology*.
- ITU-R (2014). BT.2020 - Parameter values for ultra-high definition television systems for production and international programme exchange BT Series Broadcasting service. *BT.2020 : parameter values for ultra-high definition television systems for production and international programme exchange (www.itu.int/rec/R-REC-BT.2020-2-201510-I/en)*, 1((www.itu.int/rec/R-REC-BT.2020-2-201510-I/en)).
- ITU-R (2016). Subjective assessment methods for 3D video quality. ITU-R Recommendation P.915.
- JETI (2010). *specbos 1211 - Universal Light Measurement datasheet*. JETI Technische Instrumente GmbH. JETI-1211-0110-en.
- Johnson, P. V., Kim, J., Hoffman, D. M., Vargas, A. D., and Banks, M. S. (2014). Motion artifacts on 240-Hz OLED stereoscopic 3D displays. *Journal of the Society for Information Display*, 22(8):393–403.
- Kaplanyan, A. S., Sochenov, A., Leimkuehler, T., Okunev, M., Goodall, T., and Gizem, R. (2018). Deepfovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 38(4):212:1–212:13.
- Kauvar, I., Yang, S. J., Shi, L., McDowall, I., and Wetzstein, G. (2015). Adaptive color display via perceptually-driven factored spectral projection. *ACM Transactions on Graphics*, 34(6):1–10.
- Kelly, D. H. (1959). Effects of Sharp Edges in a Flickering Field. *Journal of the Optical Society of America*, 49(7):730.
- Kelly, D. H. (1964). Sine waves and flicker fusion. *Documenta Ophthalmologica*, 18(1):16–35.
- Kelly, D. H. (1974). Spatio-temporal frequency characteristics of color-vision mechanisms\*. *Journal of the Optical Society of America*, 64(7):983.
- Kelly, D. H. (1979). Motion and vision II Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10):1340.
- Kelly, D. H. (1983). Spatiotemporal variation of chromatic and achromatic contrast thresholds. *Journal of the Optical Society of America*, 73(6):742.



- Kelly, D. H. (1984). Retinal inhomogeneity I Spatiotemporal contrast sensitivity. *Journal of the Optical Society of America A*, 1(1):107.
- Kim, K. J., Mantiuk, R., and Lee, K. H. (2013). Measurements of achromatic and chromatic contrast sensitivity functions for an extended range of adaptation luminance. In Rogowitz, B. E., Pappas, T. N., and de Ridder, H., editors, *Human Vision and Electronic Imaging XVIII*, volume 8651, page 86511A.
- Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., and Niehorster, D. (2007). What’s new in Psychtoolbox-3? A free cross-platform toolkit for psychophysics with Matlab and GNU/Octave. *Cognitive and Computational Psychophysics*, 36:1–89.
- Klompenhouwer, M. A. and Velthoven, L. J. (2004). Motion blur reduction for liquid crystal displays: motion-compensated inverse filtering. In *Visual Communications and Image Processing 2004*, volume 5308, pages 690–699. International Society for Optics and Photonics.
- Komogortsev, O. V. and Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior Research Methods*, 45(1):203–215.
- Konrad, R., Cooper, E. A., and Wetzstein, G. (2016). Novel Optical Configurations for Virtual Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 1211–1220, New York, New York, USA. ACM Press.
- Kukkonen, H., Rovamo, J., Tiippana, K., and Näsänen, R. (1993). Michelson contrast, rms contrast and energy of various spatial stimuli at threshold. *Vision Research*, 33(10):1431–1436.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86.
- Kuroki, Y., Nishi, T., Kobayashi, S., Oyaizu, H., and Yoshimura, S. (2006). 3.4: Improvement of Motion Image Quality by High Frame Rate. *SID Symposium Digest of Technical Papers*, 37(1):14.
- Kuroki, Y., Nishi, T., Kobayashi, S., Oyaizu, H., and Yoshimura, S. (2007). A psychophysical study of improvements in motion-image quality by using high frame rates. *Journal of the Society for Information Display*, 15(1):61.
- Laird, J., Rosen, M., Pelz, J., Montag, E., and Daly, S. (2006). Spatio-velocity CSF as a function of retinal velocity using unstabilized stimuli. In Rogowitz, B. E., Pappas,

- T. N., and Daly, S. J., editors, *Human Vision and Electronic Imaging XI*, volume 6057, page 605705.
- Lambrecht, C. and Verscheure, O. (1998). Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System. *Signal Processing*, pages 1–12.
- Laming, D. (2012). Weber’s law. *Inside Psychology: A Science Over 50 Years*.
- Lee, J. W., Lim, B. R., Park, R. H., Kim, J. S., and Ahn, W. (2006a). Two-stage false contour detection algorithm using re-quantization and directional contrast features and its application to adaptive false contour reduction. *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, 2006(1):377–378.
- Lee, J. W., Lim, B. R., Park, R.-H., Kim, J.-S., and Ahn, W. (2006b). Two-stage false contour detection using directional contrast and its application to adaptive false contour reduction. *IEEE Transactions on Consumer Electronics*, 52(1):179–188.
- Leigh, R. J. and Zee, D. S. (2015). *The neurology of eye movements*. OUP USA.
- Lincoln, P., Blate, A., Singh, M., Whitted, T., State, A., Lastra, A., and Fuchs, H. (2016). From Motion to Photons in 80 Microseconds: Towards Minimal Latency for Virtual and Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1367–1376.
- Lisberger, S., Evinger, C., Johanson, G., and Fuchs, A. (1981). Relationship between eye acceleration and retinal image velocity during foveal smooth pursuit in man and monkey. *Journal of Neurophysiology*, 46(2):229–249.
- Lisberger, S. G. (2010). Visual guidance of smooth-pursuit eye movements: Sensation, action, and what happens in between. *Neuron*, 66(4):477–491.
- Liu, C.-S., Bryan, R., Miki, A., Woo, J., Liu, G., and Elliott, M. (2006). Magnocellular and parvocellular visual pathways have different blood oxygen level-dependent signal time courses in human primary visual cortex. *American Journal of Neuroradiology*, 27(8):1628–1634.
- Lu, T., Pu, F., Yin, P., Pytlarz, J., Chen, T., and Husak, W. (2016). Adaptive reshaper for high dynamic range and wide color gamut video compression. *Applications of Digital Image Processing XXXIX*, 9971:99710B.
- Mackin, A., Noland, K. C., and Bull, D. R. (2016). The visibility of motion artifacts and their effect on motion quality. In *Proceedings - International Conference on Image Processing, ICIP*, volume 2016-Augus, pages 2435–2439. IEEE.

- Mäkelä, P., Rovamo, J., and Whitaker, D. (1994). Effects of luminance and external temporal noise on flicker sensitivity as a function of stimulus size at various eccentricities. *Vision Research*, 34(15):1981–1991.
- Mantiuk, R., Bazyluk, B., and Mantiuk, R. K. (2013). Gaze-driven object tracking for real time rendering. *Computer Graphics Forum*, 32(2 PART2):163–173.
- Mantiuk, R., Daly, S., and Kerofsky, L. (2008). Display adaptive tone mapping. *ACM Transactions on Graphics*, 27(3):1.
- Mantiuk, R., Daly, S. J., Myszkowski, K., and Seidel, H.-P. (2005). Predicting visible differences in high dynamic range images: model and its calibration. In Rogowitz, B. E., Pappas, T. N., and Daly, S. J., editors, *Human Vision and Electronic Imaging X*, volume 5666, page 204.
- Mantiuk, R., Kim, K. J., Rempel, A. G., and Heidrich, W. (2011). Hdr-Vdp-2. *ACM Transactions on Graphics*, 30(4):1.
- Mantiuk, R., Krawczyk, G., Myszkowski, K., and Seidel, H.-P. (2004). Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics*, 23(3):733.
- Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2006). Lossy compression of high dynamic range images and video. In Rogowitz, B. E., Pappas, T. N., and Daly, S. J., editors, *Human Vision and Electronic Imaging XI*, volume 6057, page 60570V.
- Martinez-Conde, S., Macknik, S. L., and Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature reviews neuroscience*, 5(3):229–240.
- Masia, B., Wetzstein, G., Didyk, P., and Gutierrez, D. (2013). A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics*, 37(8):1012–1038.
- McCarthy, J. D., Sasse, M. A., and Miras, D. (2004). Sharp or smooth? In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, pages 535–542, New York, New York, USA. ACM Press.
- Mikhailiuk, A., Perez-Ortiz, M., and Mantiuk, R. (2018). Psychometric scaling of TID2013 dataset. In *2018 10th International Conference on Quality of Multimedia Experience, QoMEX 2018*, pages 1–6. IEEE.
- Miller, S., Nezamabadi, M., and Daly, S. (2013). Perceptual Signal Coding for More Efficient Usage of Bit Codes. In *SMPTE Motion Imaging Journal*, volume 122, pages 52–59. IEEE.

- 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.
- Mulligan, J. B. (1993). Methods for spatiotemporal dithering. *NASA Ames Research Center*, pages 1–4.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, 1 edition.
- Navarro, F., Castillo, S., Serón, F. J., and Gutierrez, D. (2011). Perceptual considerations for motion blur rendering. *ACM Transactions on Applied Perception*, 8(3):1–15.
- Nehab, D., Sander, P. V., Lawrence, J., Tatarchuk, N., and Isidoro, J. R. (2007). Accelerating Real-Time Shading with Reverse Reprojection Caching. *SIG-GRAPH/Eurographics Sym. on Graphics Hardware*, page 11.
- Ng, K. T., Chan, S. C., and Shum, H. Y. (2005). Data compression and transmission aspects of panoramic videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(1):82–95.
- Niehorster, D. C., Siu, W. W. F., and Li, L. (2015). Manual tracking enhances smooth pursuit eye movements. *Journal of vision*, 15(15):11.
- Niklaus, S. and Liu, F. (2018). Context-Aware Synthesis for Video Frame Interpolation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1701–1710. IEEE.
- Noland, K. (2014). The application of sampling theory to television frame rate requirements. *BBC Research & Development White Paper*, 282.
- Osborne, L. C. and Lisberger, S. G. (2009). Spatial and Temporal Integration of Visual Motion Signals for Smooth Pursuit Eye Movements in Monkeys. *Journal of Neurophysiology*, 102(4):2013–2025.
- Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., and Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics*, 35(6):1–12.
- Perez-Ortiz, M. and Mantiuk, R. K. (2017). A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint*.
- Perez-Ortiz, M., Martinovic, J., Mantiuk, R., and Wuerger, S. (2019). Luminance and chromatic contrast sensitivity at high light levels. In *Perception*.

- Perrin, F. H. (1954). A Study in Binocular Flicker. *Journal of the Optical Society of America*, 44(1):60.
- Pettineo, M. (2015). Rendering The Alternate History of The Order: 1886. Presented at Advances in Real-Time Rendering in Games course at SIGGRAPH 2015.
- Plomp, R., Houtgast, T., and Steeneken, H. J. M. (1984). *Limits in perception: Essays in Honour of Maarten A. Bouman*. VSP.
- Poth, C. H., Foerster, R. M., Behler, C., Schwanecke, U., Schneider, W. X., and Botsch, M. (2018). Ultrahigh temporal resolution of visual presentation using gaming monitors and G-Sync. *Behavior Research Methods*, 50(1):26–38.
- Poynton, C. (2002). Chroma subsampling notation. *Retrieved June*, 19:3–5.
- Purcell, T. J., Buck, I., Mark, W. R., and Hanrahan, P. (2005). Ray tracing on programmable graphics hardware. In *ACM SIGGRAPH 2005 Courses*, SIGGRAPH '05, New York, NY, USA. ACM.
- Raninen, A. and Rovamo, J. (1987). Retinal ganglion-cell density and receptive-field size as determinants of photopic flicker sensitivity across the human visual field. *Journal of the Optical Society of America A*, 4(8):1620.
- Rentschler, I. and Treutwein, B. (1985). Loss of spatial phase relationships in extrafoveal vision. *Nature*, 313(6000):308–310.
- Roberts, J. and Wilkins, A. (2013a). Flicker can be perceived during saccades at frequencies in excess of 1 khz. *Lighting Research & Technology*, 45(1):124–132.
- Roberts, J. E. and Wilkins, A. J. (2013b). Flicker can be perceived during saccades at frequencies in excess of 1 kHz. *Lighting Research and Technology*, 45(1):124–132.
- Roberts, L. G. (1962). Picture Coding Using Pseudo-Random Noise. *IRE Transactions on Information Theory*, 8(2):145–154.
- Robinson, D. A. (1964). The mechanics of human saccadic eye movement. *The Journal of Physiology*, 174(2):245–264.
- Robinson, D. A. (1965). The mechanics of human smooth pursuit eye movement. *The Journal of Physiology*, 180(3):569–591.
- Robinson, D. A., Gordon, J. L., and Gordon, S. E. (1986). A model of the smooth pursuit eye movement system. *Biological cybernetics*, 55(1):43–57.

- Robson, J. G. (1966). Spatial and Temporal Contrast-Sensitivity Functions of the Visual System. *Journal of the Optical Society of America*, 56(8):1141.
- Rondao Alfaced, P., MacQ, J. F., and Verzijp, N. (2012). Interactive omnidirectional video delivery: A bandwidth-effective approach. *Bell Labs Technical Journal*, 16(4):135–147.
- Roufs, J. A. and Blommaert, F. J. (1981). Temporal impulse and step responses of the human eye obtained psychophysically by means of a drift-correcting perturbation technique. *Vision Research*, 21(8):1203–1221.
- Rutherford, A. (2003). Handbook of perception and human performance. Vol 1: Sensory processes and perception. Vol 2: Cognitive processes and performance. In *Applied Ergonomics*, volume 18, chapter 6, page 340.
- Safdar, M., Cui, G., Kim, Y. J., and Luo, M. R. (2017). Perceptually uniform color space for image signals including high dynamic range and wide gamut. *Optics Express*, 25(13):15131.
- Sajadi, B., Gopi, M., and Majumder, A. (2012). Edge-guided resolution enhancement in projectors via optical pixel sharing. *ACM Transactions on Graphics*, 31(4):1–122.
- Schaufler, G. (2002). Exploiting frame-to-frame coherence in a virtual reality system. In *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium*, pages 95–102. IEEE.
- Scherzer, D., Jeschke, S., and Wimmer, M. (2007). Pixel-correct shadow maps with temporal reprojection and shadow test confidence. *Proceedings of the 18th . . .*
- Scherzer, D., Yang, L., Mattausch, O., Nehab, D., Sander, P. V., Wimmer, M., and Eisemann, E. (2012). Temporal coherence methods in real-time rendering. *Computer Graphics Forum*, 31(8):2378–2408.
- Schied, C., Salvi, M., Kaplanyan, A., Wyman, C., Patney, A., Chaitanya, C. R. A., Burgess, J., Liu, S., Dachsbacher, C., and Lefohn, A. (2017). Spatiotemporal variance-guided filtering. *Proceedings of High Performance Graphics*, pages 1–12.
- Schütz, A. C., Braun, D. I., Kerzel, D., and Gegenfurtner, K. R. (2008). Improved visual sensitivity during smooth pursuit eye movements. *Nature Neuroscience*, 11(10):1211–1216.
- Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A., and Vorozcovs, A. (2007). High dynamic range display systems. In *ACM SIGGRAPH 2004 Papers on - SIGGRAPH '04*, page 760, New York, New York, USA. ACM Press.

- Seshadrinathan, K. and Bovik, A. (2010). Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos. *IEEE Transactions on Image Processing*, 19(2):335–350.
- SID (2012). *Information Display Measurement Standard*. Society for Information Display, 1.03b edition.
- Simonson, E. and Brozek, J. (2017). Flicker Fusion Frequency: Background and Applications. *Physiological Reviews*, 32(3):349–378.
- Smith, P. L. (1998). Bloch’s law predictions from diffusion process models of detection. *Australian Journal of Psychology*, 50(3):139–147.
- Soraci, S. and Murata-Soraci, K. (2003). *Visual Information Processing*. Perspectives on Fundamental Processes in Intellectual Functioning Series. Praeger.
- Soundararajan, R. and Bovik, A. C. (2012). Video quality assessment by reduced reference spatio-temporal entropic differencing. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(4):684–694.
- Sperling, G. and Sondhi, M. M. (1968). Model for Visual Luminance Discrimination and Flicker Detection. *Journal of the Optical Society of America*, 58(8):1133.
- Squire, L. R., Zola-Morgan, N., and Amaral, D. G. (1991). *Encyclopedia of neuroscience*. Elsevier.
- Stark, L., Young, L. R., and Vossius, G. (1962). Predictive Control of Eye Tracking Movements. *IRE Transactions on Human Factors in Electronics*, HFE-3(2):52–57.
- Stockman, A. and Sharpe, L. T. (2000). The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711–1737.
- Suh, M., Kolster, R., Sarkar, R., McCandliss, B., and Ghajar, J. (2006). Deficits in predictive smooth pursuit after mild traumatic brain injury. *Neuroscience Letters*, 401(1-2):108–113.
- Sung, K., Pearce, A., and Wang, C. (2002). Spatial-temporal antialiasing. *IEEE Transactions on Visualization and Computer Graphics*, 8(2):144–153.
- Taghavinasrabadi, A., Mahzari, A., Beshay, J. D., and Prakash, R. (2017). Adaptive 360-degree video streaming using layered video coding. In *Proceedings - IEEE Virtual Reality*, pages 347–348. IEEE.
- Takeuchi, T. and De Valois, K. K. (2005). Sharpening image motion based on the spatio-temporal characteristics of human vision. *Human Vision and Electronic Imaging X*, 5666(March 2005):83.

- Talbot, H. F. (1834). Xliv. experiments on light. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 5(29):321–334.
- Tatarchuk, N. (2006). Practical parallax occlusion mapping with approximate soft shadows for detailed surface rendering. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06, pages 81–112, New York, NY, USA. ACM.
- Templin, K., Didyk, P., Myszkowski, K., and Seidel, H.-P. (2016). Emulating displays with continuously varying frame rates. *ACM Transactions on Graphics*, 35(4):1–11.
- Terry Bahill, A. and McDonald, J. D. (1983). Smooth pursuit eye movements in response to predictable target motions. *Vision Research*, 23(12):1573–1583.
- Thaler, L., Schütz, A. C., Goodale, M. A., and Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research*, 76:31–42.
- Thibos, L. N., Still, D. L., and Bradley, A. (1996). Characterization of spatial aliasing and contrast sensitivity in peripheral vision. *Vision Research*, 36(2):249–258.
- Tourancheau, S., Callet, P. L., Barba, D., Tourancheau, S., Callet, P. L., Barba, D., Tourancheau, S., Callet, P. L., and Barba, D. (2008). Influence of motion on contrast perception : supra-threshold spatio-velocity CSF measurements To cite this version : HAL Id : hal-00252676 Influence of motion on contrast perception : supra-threshold spatio-velocity CSF measurements.
- Tourancheau, S., Le Callet, P., Brunnström, K., and André, B. (2009). Psychophysical study of LCD motion-blur perception. In Rogowitz, B. E. and Pappas, T. N., editors, *Human Vision and Electronic Imaging XIV*, volume 7240, page 724015.
- van Hateren, H. (2005). A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *Journal of Vision*, 5(4):5.
- Van Hateren, J. H. (2007). Encoding of high dynamic range video with a model of human cones. *ACM Transactions on Graphics*, 25(4):1380–1399.
- Van Nes, F. L., Koenderink, J. J., Nas, H., and Bouman, M. A. (1967). Spatiotemporal Modulation Transfer in the Human Eye. *Journal of the Optical Society of America*, 57(9):1082.
- Virsu, V. and Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research*, 37(3):475–494.
- Vlachos, A. (2015). Advanced VR rendering. In *Gdc 2015*, pages 1–3.



- Wandell, B. (1995). *Foundations of Vision*. Sinauer Associates.
- Wandell, B. A. (2011). *PeerLAB*, volume 10.
- Wang, Y., Kum, S. U., Chen, C., and Kokaram, A. (2016). A perceptual visibility metric for banding artifacts. In *Proceedings - International Conference on Image Processing, ICIP*, volume 2016-Augus, pages 2067–2071. IEEE.
- Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multi-scale structural similarity for image quality assessment. *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, 2:1398–1402.
- Watson, A. B. (1990). Models in Early Vision. In *Human Performance Models for Computer-Aided Engineering*, pages 61–74. Elsevier.
- 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. (2015). High Frame Rates and Human Vision: A View through the Window of Visibility. *SMPTE Motion Imaging Journal*, 122(2):18–32.
- Watson, A. B. and Ahumada, A. J. (2011). Blur clarified: A review and synthesis of blur discrimination. *Journal of Vision*, 11(5):10–10.
- Watson, A. B. and Ahumada, A. J. (2017). The pyramid of visibility. In *Electronic Imaging*, volume 2016, pages 1–6.
- Watson, A. B., Ahumada, A. J., and 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.
- Watson, A. B. and Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2):113–120.
- Watson, A. B. and Solomon, J. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, 14(9):2379–2391.
- Weibull, W. (1951). Wide applicability. *Journal of applied mechanics*, 103(730):293–297.
- Westerink, J. H. and Teunissen, C. (1990). Perceived sharpness in moving images. In *Human Vision and Electronic Imaging: Models, Methods, and Applications*, volume 1249, pages 78–87. International Society for Optics and Photonics.

- Wilson, H. R. and Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research*, 19(1):19–32.
- Wolski, K., Giunchi, D., Ye, N., Didyk, P., Myszkowski, K., Mantiuk, R., Seidel, H.-P., Steed, A., and Mantiuk, R. K. (2018). Dataset and Metrics for Predicting Local Visible Differences. *ACM Transactions on Graphics*, 37(5):1–14.
- Wuerger, S., Ashraf, M., Kim, M., Martinovic, J., Pérez-Ortiz, M., and Mantiuk, R. K. (2019). Spatio-chromatic contrast sensitivity under mesopic and photopic light levels. *Journal of Vision*, in print.
- Yang, L., Nehab, D., Sander, P. V., Sitthi-amorn, P., Lawrence, J., and Hoppe, H. (2009). Amortized supersampling. *ACM Transactions on Graphics*, 28(5):1.
- Ye, N., Wolski, K., and Mantiuk, R. K. (2019). Predicting visible image differences under varying display brightness and viewing distance. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (CVPR 2019)*.
- Ye, P. and Doermann, D. (2014). Active sampling for subjective image quality assessment. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 4249–4256.
- Yeshurun, Y. and Levy, L. (2003). Transient spatial attention degrades temporal resolution. *Psychological Science*, 14(3):225–231.
- Yu, M., Lakshman, H., and Girod, B. (2015). A framework to evaluate omnidirectional video coding schemes. In *Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2015*, pages 31–36. IEEE.
- Zele, A. J. and Cao, D. (2015). Vision under mesopic and scotopic illumination. *Frontiers in psychology*, 5:1594.
- Zerman, E., Hulusic, V., Valenzise, G., Mantiuk, R., and Dufaux, F. (2017). Effect of color space on high dynamic range video compression performance. In *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*, pages 1–6. IEEE.
- Zerman, E., Hulusic, V., Valenzise, G., Mantiuk, R., and Dufaux, F. (2018). The relation between mos and pairwise comparisons and the importance of cross-content comparisons.

---

---

# APPENDIX A

---

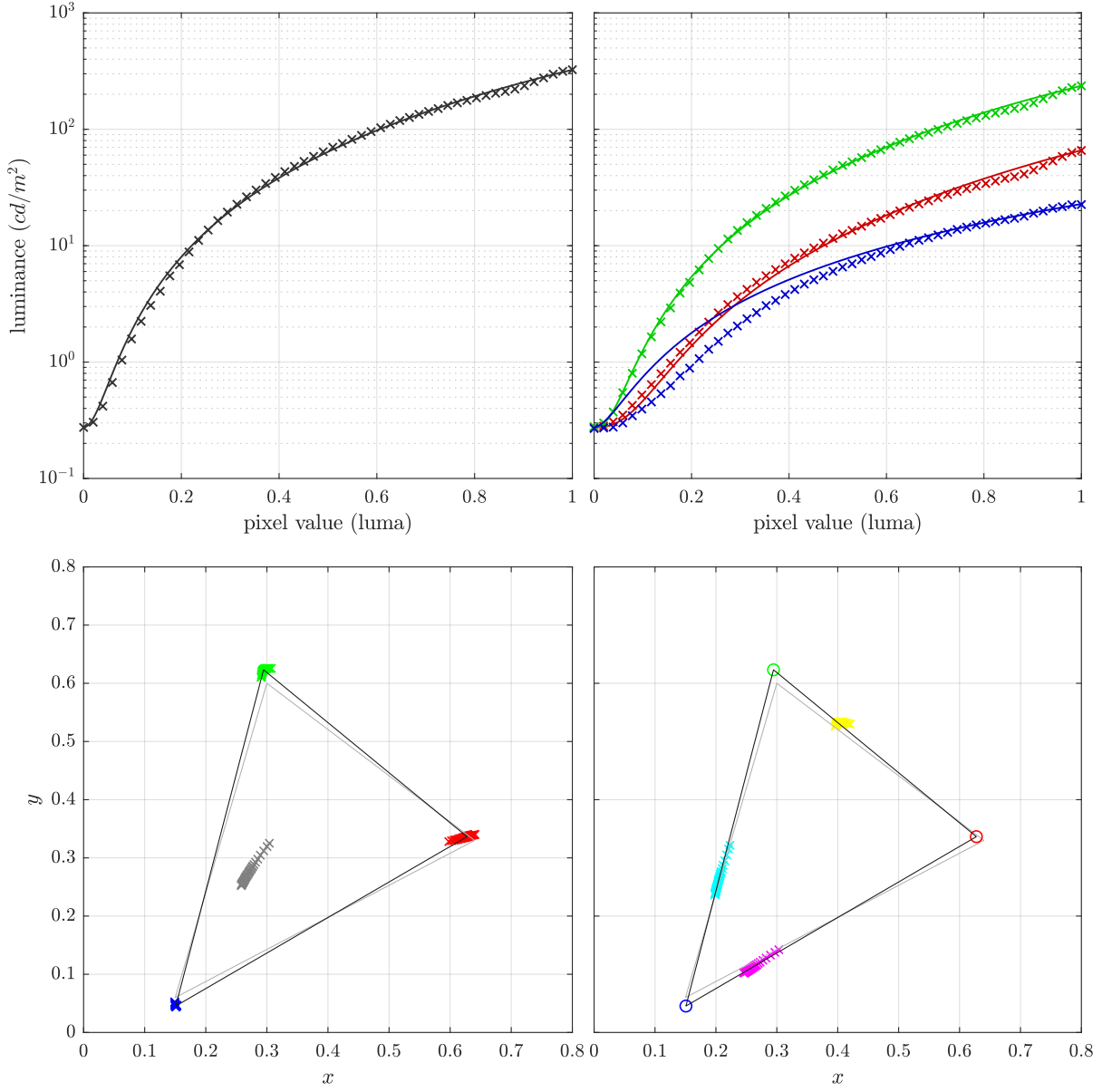
## DISPLAY PROFILES

This appendix section contains detailed results of the display measurement results including the least-square-fit gamma-offset-gain model parameters.

## A.1 Dell Inspiron 17R 7720 3D

Peak luminance:  $324.90 \text{ cd/m}^2$

dynamic range: 1184:1



**profile**

$$\gamma = (2.54, 2.38, 1.68),$$

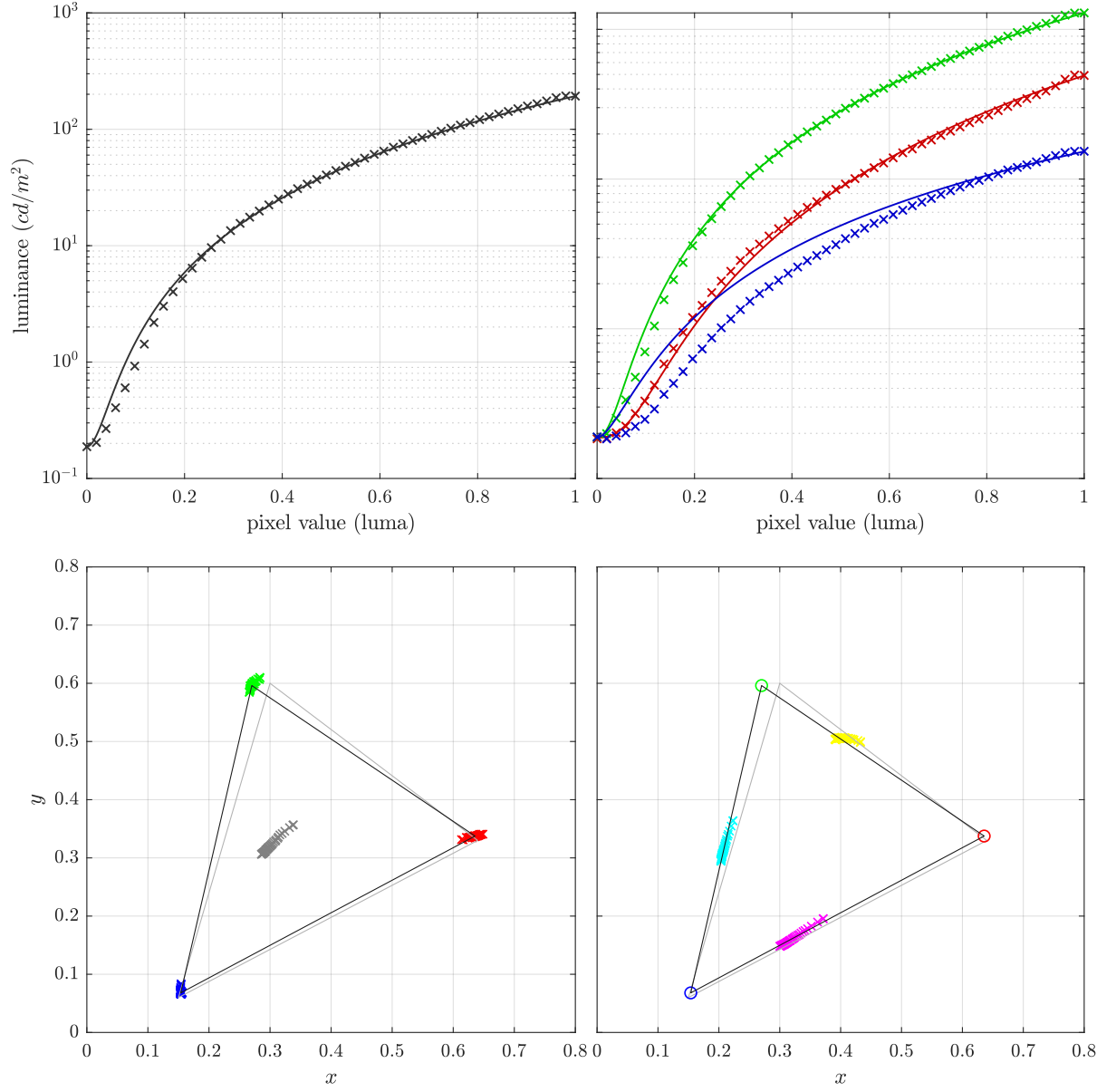
$$\text{black level: } b_{XYZ} = (0.2829, 0.2742, 0.5714),$$

$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0109 & -0.0030 & 0.0001 \\ -0.0052 & 0.0057 & -0.0004 \\ -0.0017 & 0.0002 & 0.0029 \end{bmatrix}$$

## A.2 Samsung SyncMaster

Peak luminance:  $192.80 \text{ cd/m}^2$

dynamic range: 1029:1



**profile**

$$\gamma = (2.51, 2.18, 1.68),$$

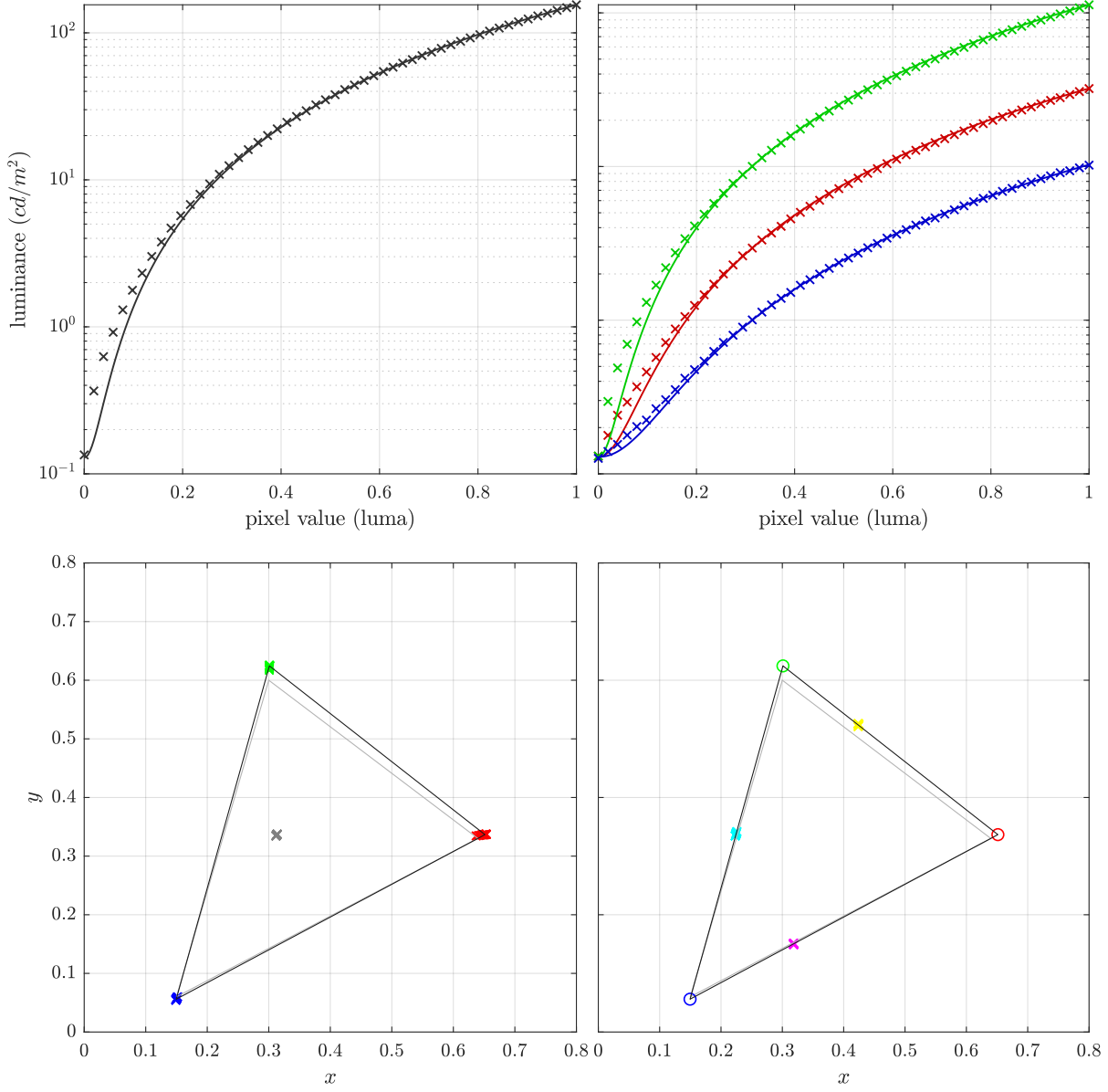
$$\text{black level: } b_{XYZ} = (0.1933, 0.1873, 0.3158),$$

$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0139 & -0.0054 & 0.0007 \\ -0.0061 & 0.0103 & -0.0016 \\ -0.0022 & -0.0000 & 0.0072 \end{bmatrix}$$

## A.3 ASUS PG279Q

Peak luminance:  $155.20\text{cd}/\text{m}^2$

dynamic range: 1154:1



**profile**

$$\gamma = (2.10, 2.10, 2.12),$$

$$\text{black level: } b_{XYZ} = (0.1266, 0.1289, 0.2192),$$

$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0211 & -0.0061 & 0.0005 \\ -0.0098 & 0.0117 & -0.0010 \\ -0.0033 & 0.0003 & 0.0066 \end{bmatrix}$$

### A.3.1 High-refresh-rate model parameters

#### 90 Hz

$$\gamma = (2.05, 2.05, 2.04),$$

$$\text{black level: } b_{XYZ} = (0.1237, 0.1255, 0.2201),$$

$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0209 & -0.0059 & 0.0004 \\ -0.0096 & 0.0115 & -0.0009 \\ -0.0033 & 0.0003 & 0.0063 \end{bmatrix},$$

$$p_{rgb} = \begin{bmatrix} 0.1888 & -0.2976 & 0.4957 & -0.0315 \\ 0.2326 & -0.3554 & 0.4821 & -0.0282 \\ 0.2365 & -0.3754 & 0.4900 & -0.0326 \end{bmatrix}$$

#### 120 Hz

$$\gamma = (2.06, 2.05, 2.07),$$

$$\text{black level: } b_{XYZ} = (0.1237, 0.1258, 0.2225),$$

$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0209 & -0.0059 & 0.0004 \\ -0.0096 & 0.0115 & -0.0009 \\ -0.0033 & 0.0003 & 0.0063 \end{bmatrix},$$

$$p_{rgb} = \begin{bmatrix} 0.1965 & -0.3145 & 0.4954 & -0.0282 \\ 0.2281 & -0.3679 & 0.4937 & -0.0313 \\ 0.2409 & -0.3959 & 0.4939 & -0.0294 \end{bmatrix}$$

#### 165 Hz

$$\gamma = (2.08, 2.07, 2.08),$$

$$\text{black level: } b_{XYZ} = (0.1307, 0.1285, 0.2214),$$

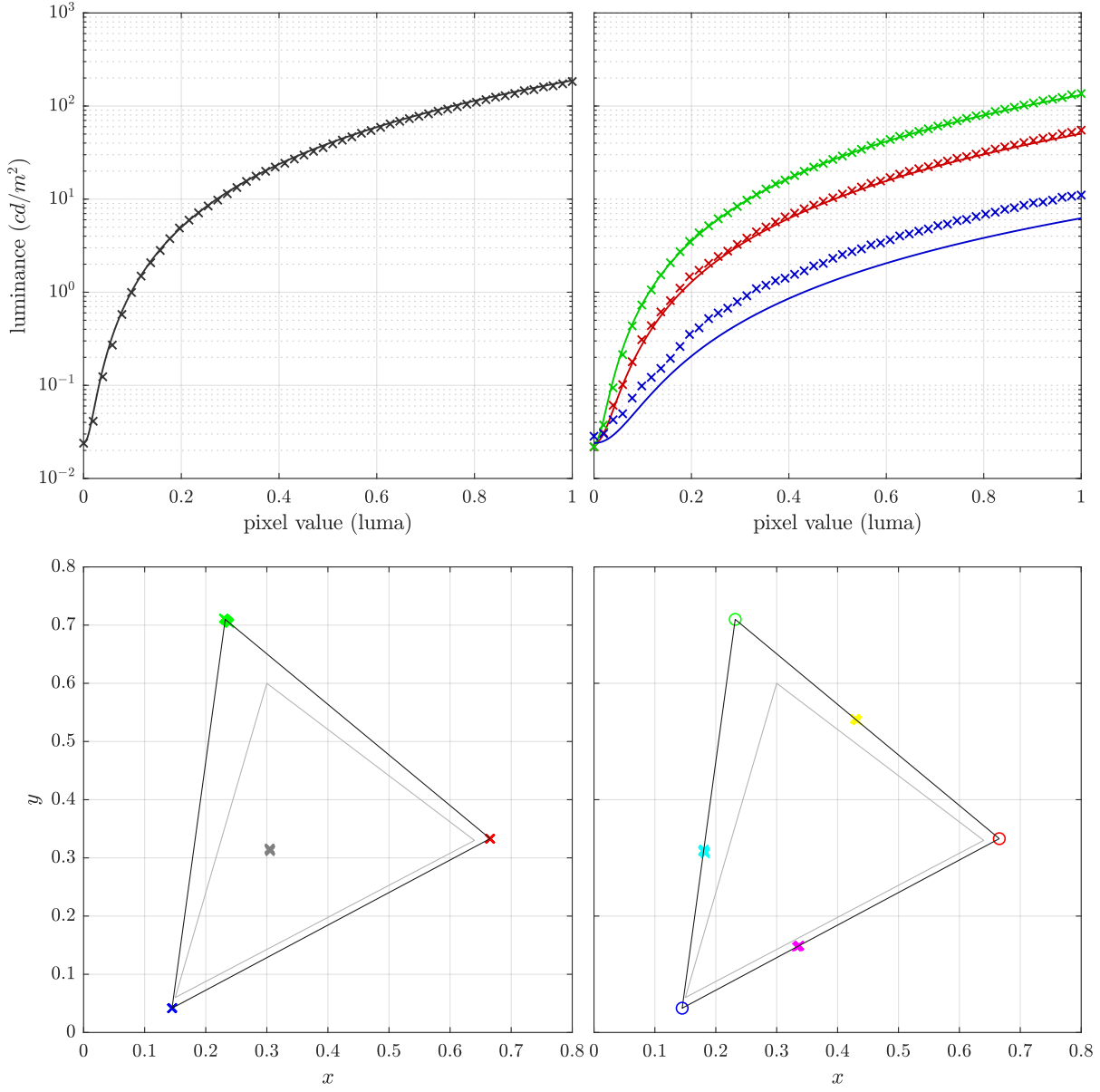
$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0208 & -0.0059 & 0.0004 \\ -0.0095 & 0.0114 & -0.0010 \\ -0.0032 & 0.0003 & 0.0063 \end{bmatrix},$$

$$p_{rgb} = \begin{bmatrix} 0.2054 & -0.3433 & 0.4986 & -0.0259 \\ 0.2372 & -0.3863 & 0.4932 & -0.0278 \\ 0.2601 & -0.4331 & 0.4908 & -0.0253 \end{bmatrix}$$

## A.4 HTC Vive

Peak luminance:  $183.32 \text{ cd/m}^2$

dynamic range: 7667:1



**profile**

$$\gamma = (2.29, 2.25, 2.19),$$

$$\text{black level: } b_{XYZ} = (0.0313, 0.0240, 0.0125),$$

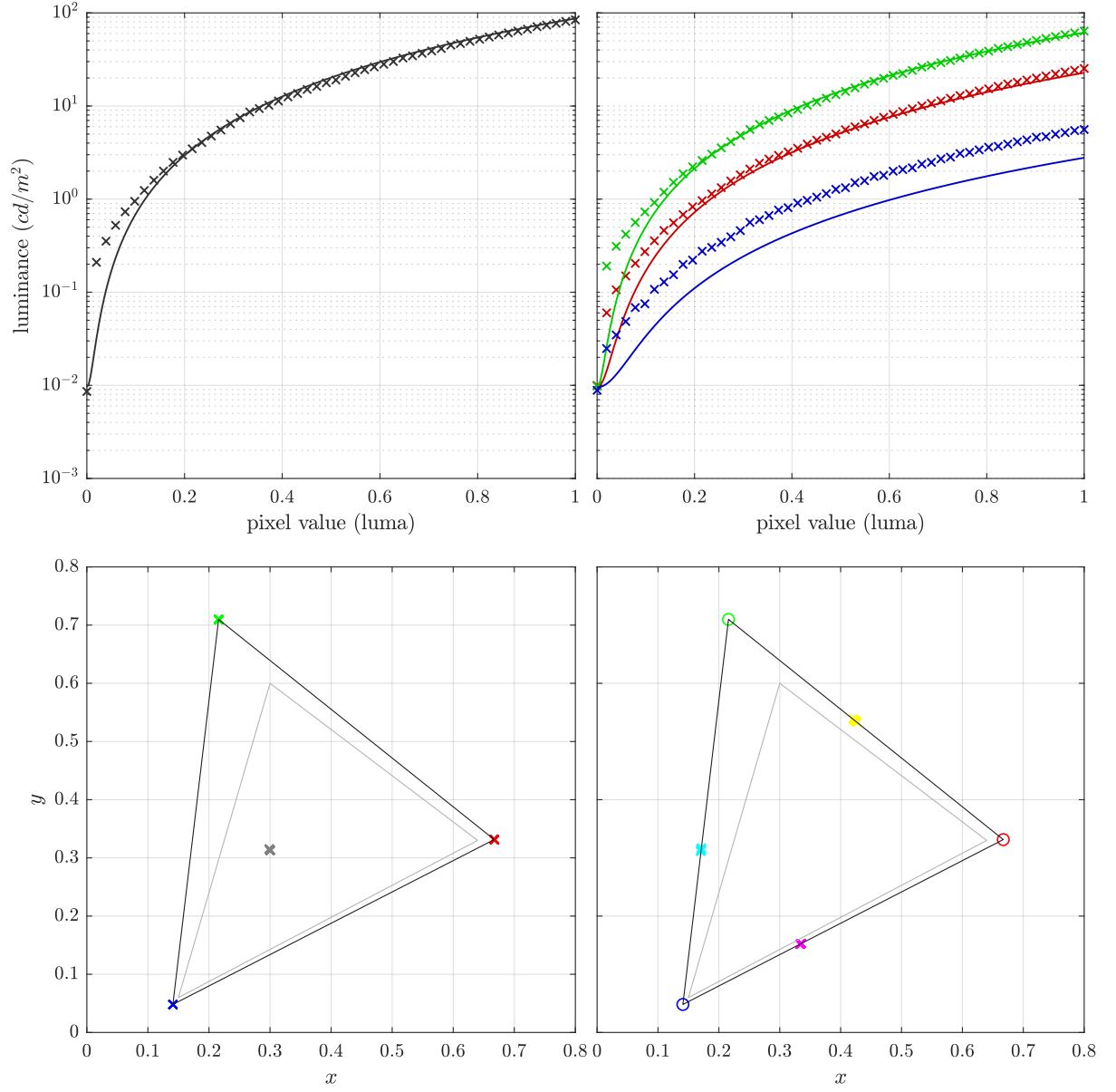
$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0109 & -0.0042 & 0.0003 \\ -0.0032 & 0.0089 & -0.0004 \\ -0.0017 & 0.0004 & 0.0046 \end{bmatrix}$$



## A.5 Oculus Rift

Peak luminance:  $84.26 \text{ cd/m}^2$

dynamic range: 9810:1



**profile**

$$\gamma = (2.15, 2.09, 2.06),$$

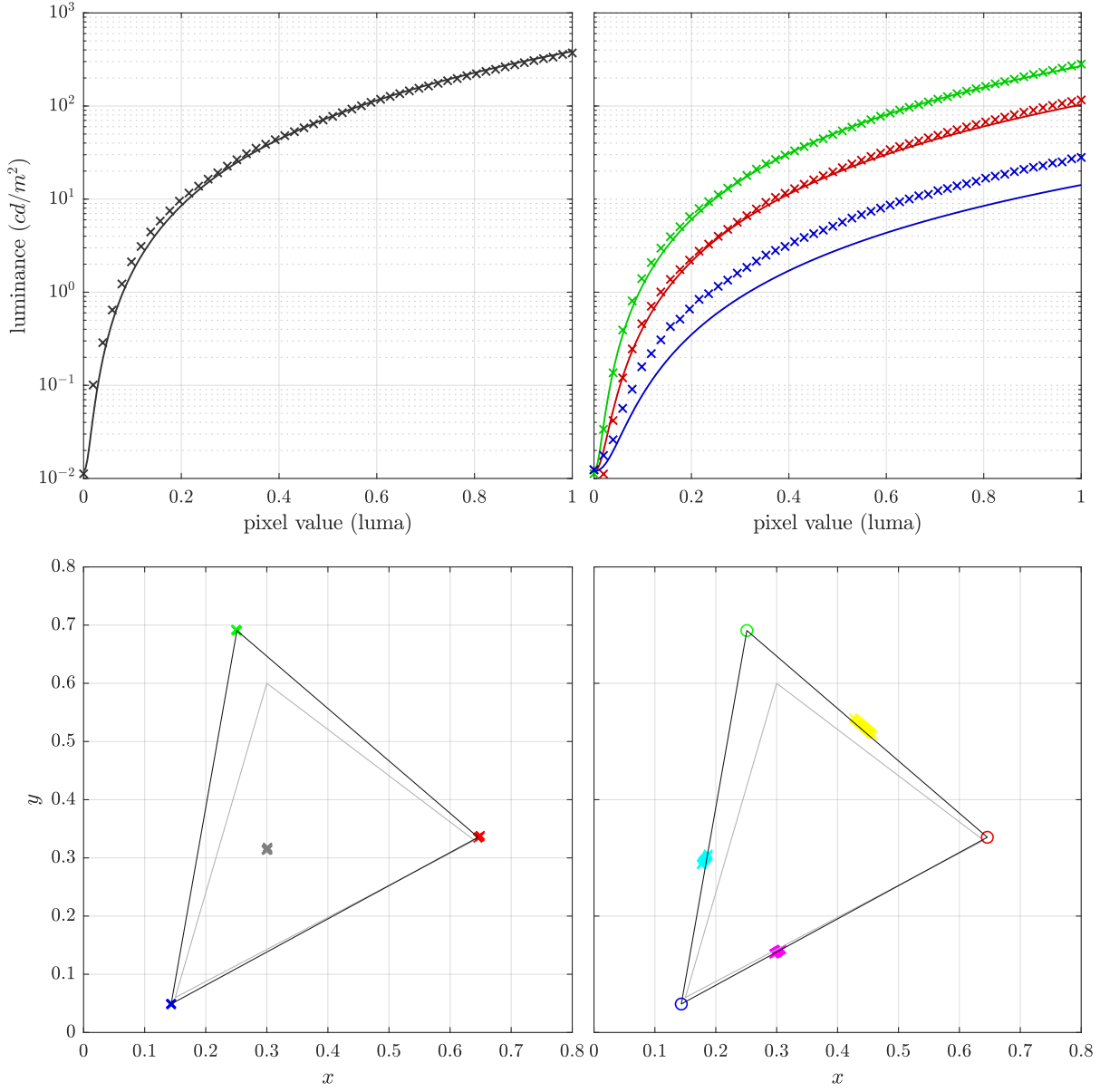
$$\text{black level: } b_{XYZ} = (0.0079, 0.0095, 0.0073),$$

$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0232 & -0.0086 & 0.0005 \\ -0.0064 & 0.0187 & -0.0014 \\ -0.0034 & 0.0008 & 0.0105 \end{bmatrix}$$

## A.6 Huawei Mate Pro 9 – normal mode

Peak luminance:  $371.40 \text{ cd/m}^2$

dynamic range: 33102:1



**profile**

$$\gamma = (2.41, 2.36, 2.32),$$

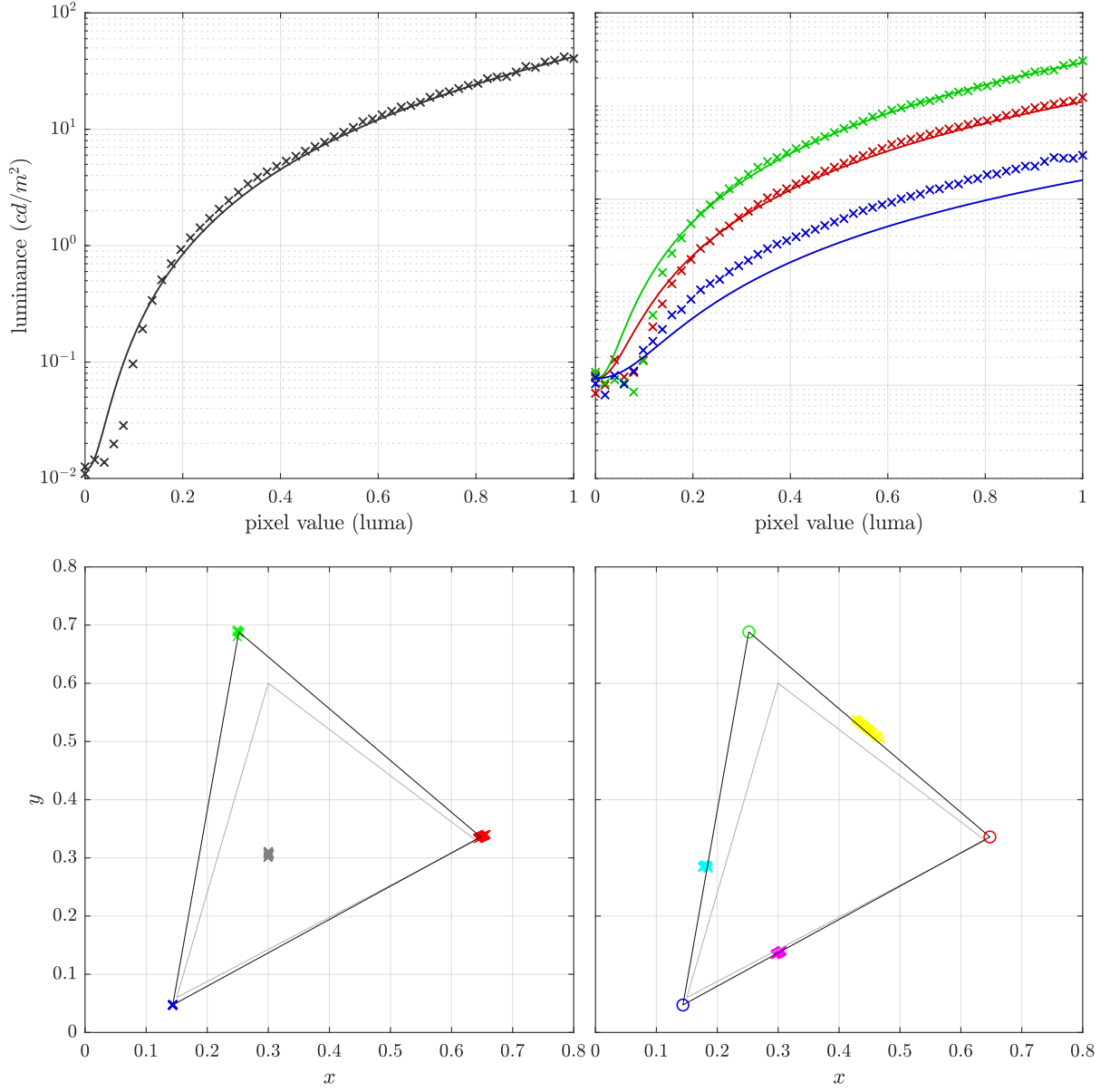
$$\text{black level: } b_{XYZ} = (0.0126, 0.0120, 0.0247),$$

$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0056 & -0.0022 & 0.0001 \\ -0.0018 & 0.0044 & -0.0002 \\ -0.0008 & 0.0002 & 0.0022 \end{bmatrix}$$

## A.7 Huawei Mate Pro 9 – VR mode

Peak luminance:  $40.50 \text{ cd/m}^2$

dynamic range: 3695:1



**profile**

$$\gamma = (2.39, 2.46, 2.28),$$

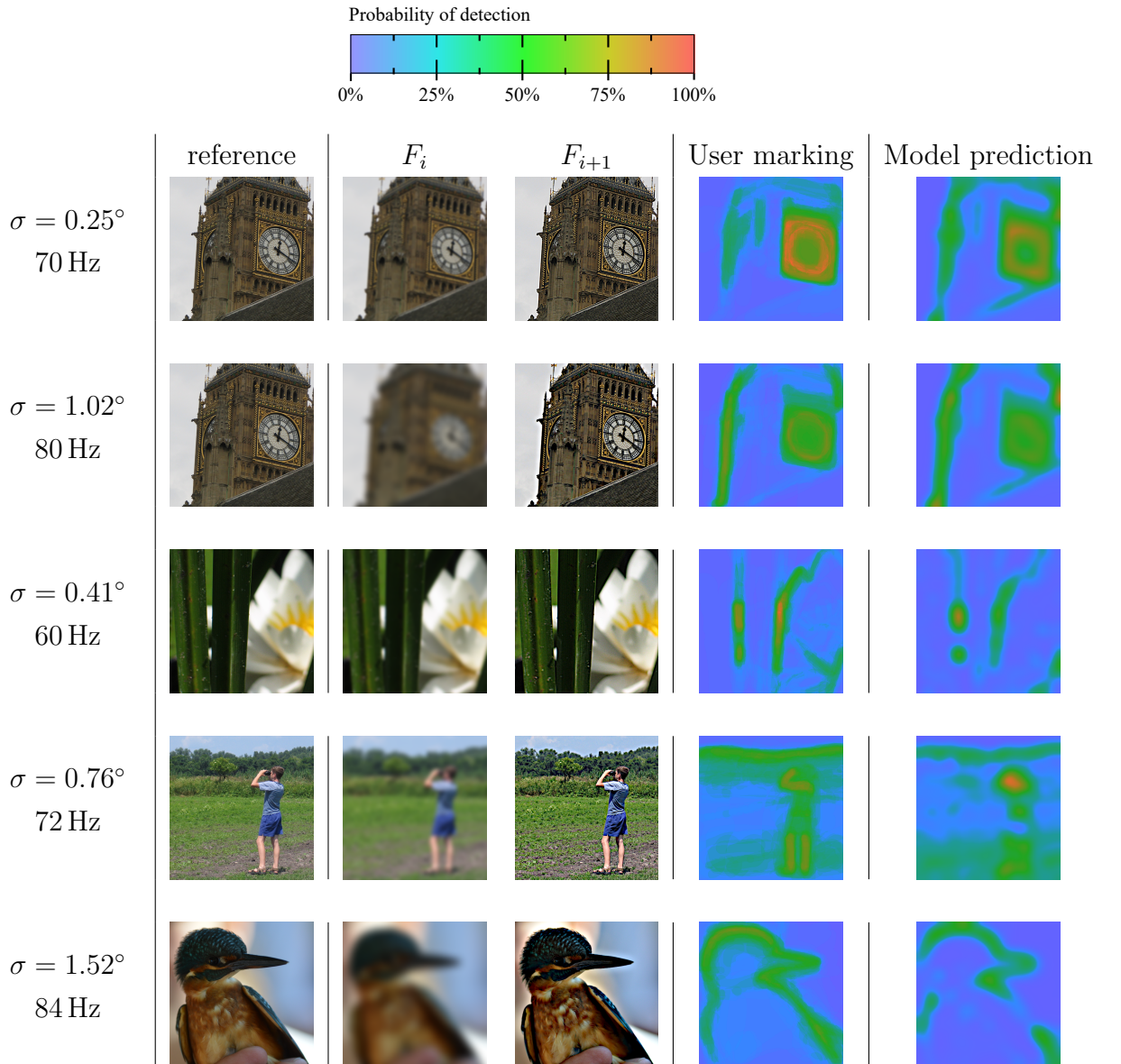
$$\text{black level: } b_{XYZ} = (0.0178, 0.0119, 0.0144),$$

$$M_{XYZ \rightarrow RGB} = \begin{bmatrix} 0.0527 & -0.0200 & 0.0008 \\ -0.0174 & 0.0409 & -0.0014 \\ -0.0077 & 0.0018 & 0.0199 \end{bmatrix}$$

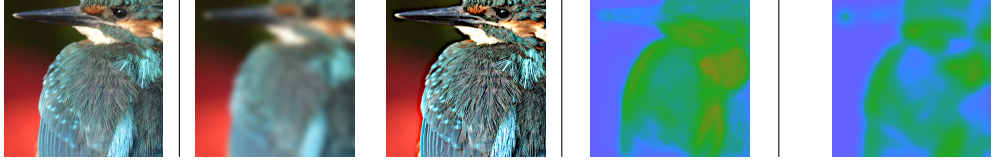


# APPENDIX B

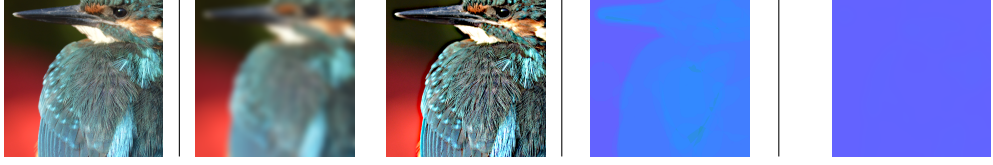
## FLICKER MARKING STIMULI



$\sigma = 1.02^\circ$   
80 Hz



$\sigma = 1.02^\circ$   
120 Hz



$\sigma = 0.10^\circ$   
62 Hz



$\sigma = 2.03^\circ$   
90 Hz



$\sigma = 1.02^\circ$   
85 Hz



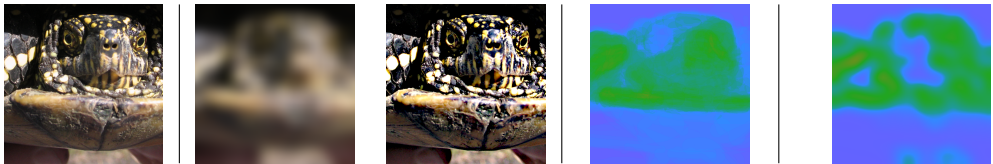
$\sigma = 0.25^\circ$   
70 Hz






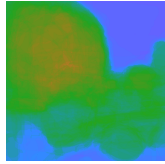
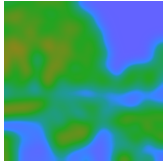

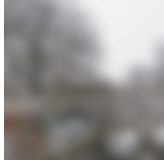

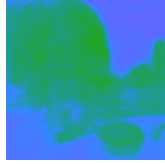
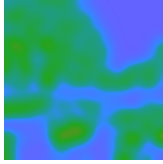



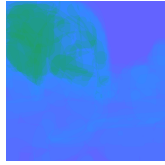
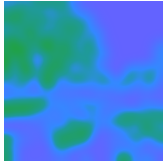

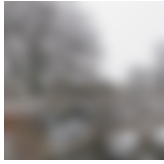

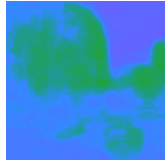
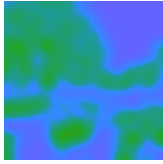
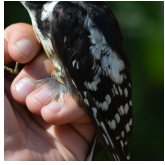


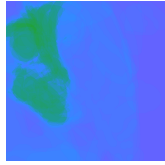
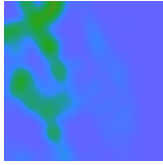
$\sigma = 0.46^\circ$   
70 Hz



$\sigma = 2.03^\circ$   
90 Hz





$\sigma = 1.02^\circ$ 80 Hz					
$\sigma = 2.03^\circ$ 88 Hz					
$\sigma = 1.02^\circ$ 90 Hz					
$\sigma = 2.03^\circ$ 90 Hz					
$\sigma = 1.52^\circ$ 90 Hz					





---

# APPENDIX C

---

## MOTION QUALITY MODEL PREDICTIONS

This appendix section is a qualitative extension to the ablation study in Section 8.3.5. The proposed complete blur-judder model (red colour in the plots) provides a good fit within the target refresh rate range. Furthermore, unlike alternative models, predictions below and above the target refresh rate range are also plausible.

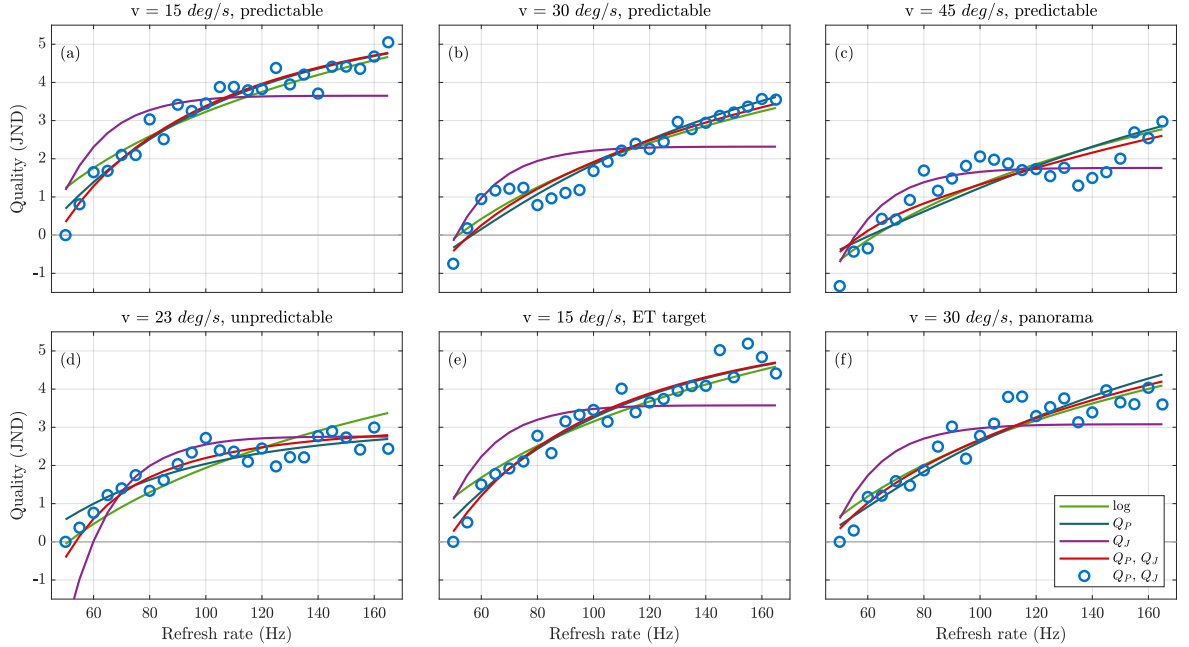


Figure C.1: Predictions of different model versions for the target refresh rate range (from 50 Hz to 165 Hz). With the exception of the judder-only model ( $Q_J$ ), all models provide reasonable predictions.

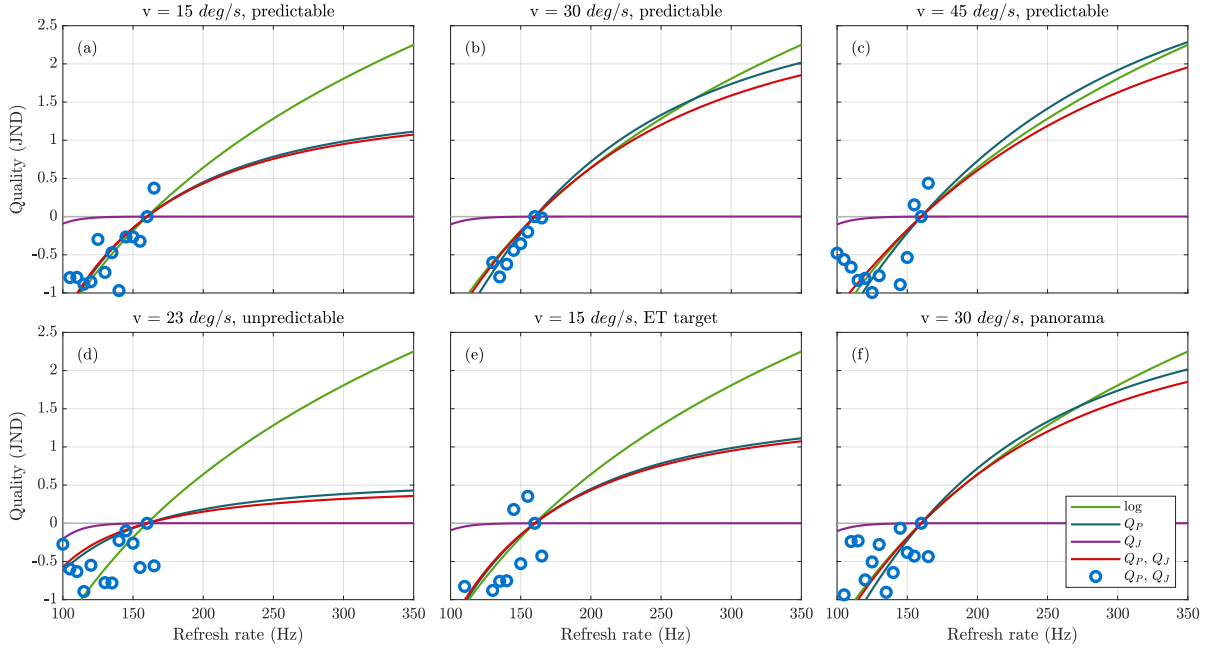


Figure C.2: Predictions of different model versions above the target refresh rate range ( $>165$  Hz). The complete proposed model (red) provides the most reasonable predictions; the log-model (green) is inconsistent with the visual system's diminishing ability to differentiate between high refresh rates; the judder model (purple) is inconsistent with existing measurements which show that humans can differentiate between refresh rates above 150 Hz.

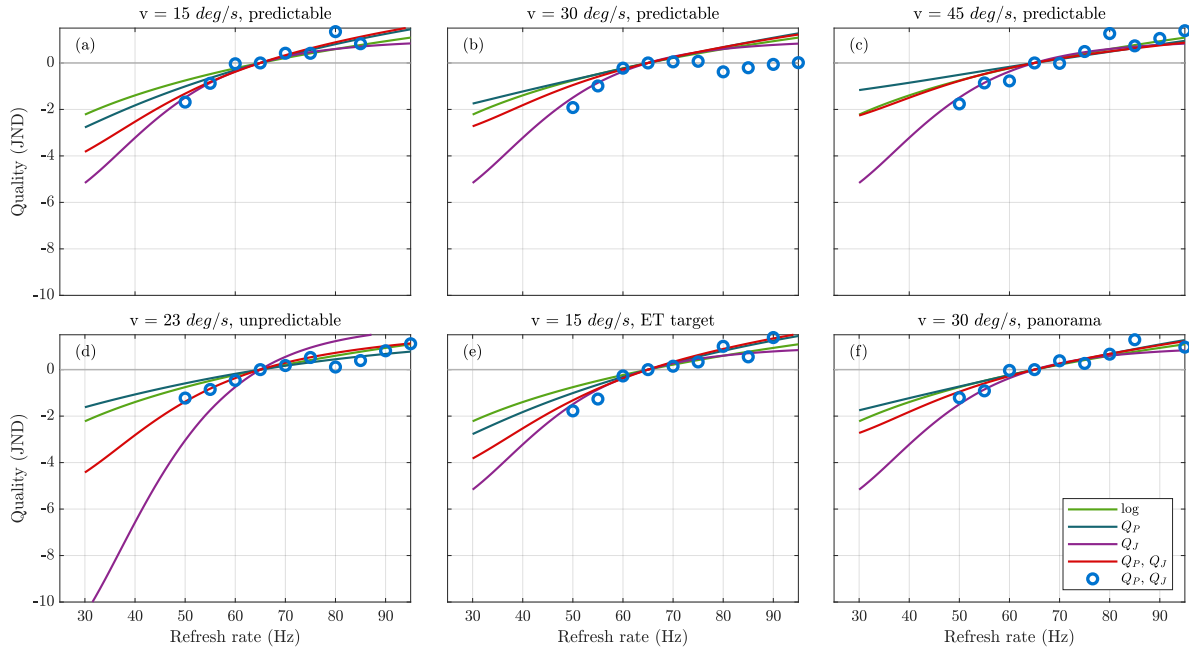


Figure C.3: Predictions of different model versions below the target refresh rate range ( $<50$  Hz). With such low refresh rates, the quality curve is expected to be reasonably steep. The judder-only model provides perhaps the most intuitive results, with the complete proposed model (red) coming second best. For more accurate predictions, flicker artefacts would also need to be considered.