University of California
Santa Barbara

Virtual Reality Simulations for the Advancement of Visual Prosthetics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Dynamical Neuroscience

by

Justin Michael Kasowski

Committee in charge:

Professor Michael Beyeler, Chair
Professor Miguel Eckstein
Professor Mary Hegarty

December 2023

The Dissertation of Justin Michael Kasowski is approved.

Professor Miguel Eckstein

Professor Mary Hegarty

Professor Michael Beyeler, Committee Chair

October 2023

Virtual Reality Simulations for the Advancement of Visual Prosthetics

by

Justin Michael Kasowski

I dedicate the following to my professors at Santa Barbara City College. For your inspirational teaching which started me on this path and has motivated me to inspire future generations in the same way, thank you.

Acknowledgements

First, I need to acknowledge my mentor Dr. Michael Beyeler. He took a risk accepting a student without any formal training in computer science and showed extreme patience while I learned a new field. He fostered a welcoming environment with a strong focus on respect and mental well-being. He inspired productivity through his own accomplishments instead of unrealistic demands, provided guidance when it was needed, and freedom when it was not. I genuinely could not have asked for a more fitting mentor and will never be able to put into words the credit he deserves.

Second, and equally important, is my wife. She has been an incredible support and a more-than-willing guinea pig in all of my experiments. She has proofread more academic papers than she ever should have and was an incredible color-checker to a colorblind scholar. I probably could have made it through this without her, but I wouldn't want to.

I would also like to acknowledge everyone else who made this possible. Dr. Elizabeth Chrastil, for seeing my potential when others may have overlooked it, for providing guidance and inspiration to a budding scholar, and most of all, for her patience and understanding of the demands in graduate school. My committee members, professors Miguel Eckstein and Mary Hegarty, for allowing me to add chaos to their already impacted schedule and holding me to their rigorous academic standards. Lastly, lab members past and present, who made spending countless hours on computers a lot less miserable than it should have been.

There are too many others to acknowledge directly, but my sincerest gratitude to my friends, family members, and colleagues who supported me during these endeavors.

Curriculum Vitæ

Justin Michael Kasowski

Education

2023	Ph.D. in Dynamical Neuroscience (Expected), University of California, Santa Barbara.
2018	B.S. in Cell and Developmental Biology, University of California, Santa Barbara (2016-2018)

Publications

J. Kasowski, B.A. Johnson, R. Neydavood, A. Akkaraju, and M. Beyeler (2023). "A systematic review of extended reality (XR) for understanding and augmenting vision loss". Journal of Vision, Special Edition on Mixed Reality, 23(5):5, 1–24, https://doi.org/10.1167/jov.23.5.5.

J. Kasowski and M. Beyeler (2022). "Immersive Virtual Reality Simulations of Bionic Vision". Augmented Humans, 2022. ACM, Kashiwa, Chiba Japan, 82-93. https://doi.org/10.1145/3519391.3522752

J. Kasowski, N. Wu, and M. Beyeler. "Towards Immersive Virtual Reality Simulations of Bionic Vision". Augmented Humans Conference 2021 (AHs'21). Association for Computing Machinery, New York, NY, USA, 313–315. https://doi.org/10.1145/3458709.3459003

Presentations

J. Kasowski, M. Beyeler. "SimpleXR - An open-source Unity toolbox for simplifying vision research using augmented and virtual reality". Platform presentation, Vision Science Society, 2023.

J. Kasowski, M. Beyeler. "Augmented Reality Simulation of Bionic Vision". Demo Night, VSS 2023.

J. Kasowski, "Immersive Virtual Reality Simulations of Bionic Vision". Platform presentation, Eye and the Chip, 2021.

J. Kasowski, E. Chrastil. "Assessment of Individual Differences in Navigation by Diffusion MRI Connectometry." Poster, Society for Neuroscience Annual Meeting, 2019.

Abstract

Virtual Reality Simulations for the Advancement of Visual Prosthetics

by

Justin Michael Kasowski

The fields of visual prosthetics and virtual reality (VR) are intersecting in exciting ways. Designed to restore a rudimentary form of vision to people living with profound blindness, visual prostheses electrically stimulate surviving cells in the visual pathway to evoke visual percepts. Like VR headsets, these devices commonly use a head-mounted camera to capture visual data, updating the view as the user shifts position. Despite the growing use of VR headsets to simulate what people "see" using visual prostheses, most previous simulations lack biological realism or do not consider the way a prosthesis user would use head and eye movements to sample the scene.

To address these challenges, I developed BionicVisionXR, an open-source VR toolbox for simulated prosthetic vision that uses a neurophysiologically inspired and psychophysically validated computational model to allow sighted participants to 'see through the eyes' of a prosthesis user. First, to demonstrate its utility, I systematically evaluated the effect of clinically reported perceptual distortions on performance in letter recognition and immersive obstacle avoidance tasks. Second, I enriched our simulations with gaze contingency and temporal effects to capture often neglected simulation parameters that may affect the quality of vision provided by existing devices. Third, to guide the development of next-generation devices, I propose a way of decomposing the scene into meaningful parts using edge detection and semantic segmentation. My results demonstrate the importance of choosing an appropriate level of immersion and phosphene model complexity. Furthermore, using a combination of computational modeling and behav-

ioral testing, I was able to identify electrode stimulation strategies that may improve the quality of vision provided by retinal implants.

This work has the potential to 1) further our understanding of the qualitative experience associated with different visual prosthetics, 2) provide realistic expectations of prosthetic performance for patients, doctors, manufacturers, and regulatory bodies, and 3) accelerate the prototyping of new devices that may one day restore useful vision to people living with profound blindness.

Contents

Chapter 1

Introduction

1.1 Motivation

The World Health Organization predicts approximately 114.6 million people will be living with incurable blindness by the year 2050, and 587.6 million people will be affected by severe visual impairment (Bourne et al., 2017). While some of these cases can be addressed with medication or surgical interventions, many who experience profound retinal degeneration or damage to the optic nerve or cortex have no available effective treatments. For such individuals, an electronic visual prosthesis, often termed a "bionic eye", could be the sole solution (Fernandez, 2018). Similar in principle to cochlear implants, these devices use electrical impulses to activate the remaining functional cells in the visual pathway, resulting in visual sensations known as "phosphenes". Although the phosphenes induced by current prosthetics do enhance users' capabilities in discerning high-contrast items and conducting basic navigation tasks (Ayton et al., 2020), they cannot yet replicate the clarity of natural sight. Despite the transformative potential of bionic eyes for those with irreversible blindness, globally, only about 500 retinal prostheses have been installed. To explore functional

restoration and experiment with diverse implant architectures, scientists have been crafting virtual reality (VR) models anchored on simulated prosthetic vision (SPV). The conventional approach engages sighted subjects experiencing vision solely through phosphenes rendered on a head-mounted display (HMD). This offers these participants a firsthand perspective of a bionic eye user's visual experience, factoring in their head or eye movements while navigating a digital realm (Kasowski et al., 2021).

XR platforms have been instrumental in simulating the visual experience offered by neuroprosthetic devices, providing valuable insights into their functional outcomes. Presently, several research initiatives leverage XR tools to model, analyze, and enhance the outcomes of visual neuroprostheses. While the simulations offer critical insights, they often diverge from the actual experiences of prosthetic recipients. The visual outputs generated by bionic devices, owing to the intricate neural-implant interface, can vary significantly from natural vision (Erickson-Davis and Korzybska, 2021). Additionally, many simulated prosthetic vision (SPV) experiments merely showcase stimuli on a PC screen or within an HMD, often disregarding essential elements like eye movements, head motions, or walking (Kasowski et al., 2023). This leads to a low level of immersion (Kardong-Edgren et al., 2019; Pasch et al., 2009) which refers to technical manipulations that separate the existence of the physical world from the virtual world (Miller and Bugnariu, 2016). Interactions with the real world, such as using tangible controls or hearing ambient sounds, can diminish this immersive experience. Yet, the significance of immersion in SPV behavioral tasks remained a mystery, as no previous study had evaluated if behavioral outcomes remain consistent across tasks performed on a monitor and those on an HMD.

Moreover, the majority of current prosthetic solutions offer a quite restricted field of view (FOV). For example, the visual field produced by the widely-used Argus II implant (Luo and da Cruz, 2016b) is constrained to a mere 10×20 degrees of visual

perspective. This limitation necessitates users to make deliberate head movements to gather visual information from their surroundings (Erickson-Davis and Korzybska, 2021). The advent of immersive VR in academic research equips scholars with the tools to authentically emulate this scenario.

After systematically reviewing over 11,000 search results, I found most SPV recent studies relied on phosphene models with a low level of biological realism or immersion (Kasowski et al., 2023). Thus, it remained uncertain how the outcomes of previous SPV experiments would apply to real prosthesis users.

As the field of XR continues to grow, with advancements in hardware and software, the tools at our disposal become more refined, leading to more accurate and insightful simulations. However, there remains a critical need to bridge the gap between simulation and reality – ensuring that our models reflect the lived experiences of actual prosthetic recipients. This dissertation, seeks to explore the use of XR technologies for simulating visual neuroprostheses by evaluating previous works and introducing a new toolbox for studying simulated prosthetic vision in XR.

1.2 Aims and Objectives

The main goal of this thesis is to utilize virtual reality simulations to better understand, and possibly enhance, prosthetic vision. This work requires an immersive, spatio-temporally accurate, and biologically real model of bionic vision. Previous models of prosthetic vision are often simplified or presented as static images on a computer monitor.

This research encompasses the design of computational models that are informed by neuroanatomical and neurophysiological evidence about the early visual system, along with the development of open-source software libraries that provide the means to study

these models, conduct visual psychophysics studies, and evaluate theories in a realistic virtual environment. As a result of these efforts, the contributions of this thesis span multiple disciplines including computational neuroscience, computer science, vision science, and psychology.

In specific, the contributions of this thesis include:

1. BionicVisionXR, an open-source VR toolbox for simulated prosthetic vision that allows sighted participants to "see through the eyes" of a prosthesis user. Importantly, the toolbox uses an established and psychophysically validated computational model of bionic vision Beyeler et al. (2019b) to generate realistic SPV predictions.

2. The use of BionicVisionXR to systematically evaluate how different display types (HMD or monitor) and clinically reported spatial distortions affect performance across a variety of behavioral tasks. These factors have largely been ignored, with the majority of studies treating each electrode as if it were a bright circle. This kind of model is referred to as the 'scoreboard model', because like a sports stadium scoreboard, the combination of these circles can be used to form words or images. While electrical implants could potentially operate in a similar way, currently there are no devices able to achieve this in real life. In reality, there are spatial distortions, along with other often overlooked factors like a head fixed camera that may not respond to eye movements. Not including these factors may drastically change performance, but to the best of our knowledge, this is the first SPV study to address these factors. We use a within-subjects design to allow for a direct comparison between display types of the same tasks, and have our subjects perform the task on a broad range of spatial distortion values.

3. The use of BionicVisionXR to systematically evaluate how temporal effects factor

into behavioral performance. In neuroprostheses, the electrodes are not all continuously active. Instead, they are activated in smaller groups in a process known as rasterization. This limits the overall charge being injected, which stalls neural desensitization and can conserve the device's battery. However, it is not clear whether there is an optimal pattern as developers are likely to either use seemingly random grouping (Second Sight, 2013), or basic top down/left to right rastering. To our knowledge, the second study in this work is the first of its kind to study the effect of electrode rasterization in VR.

4. The use of BionicVisionXR to discover advanced stimulation strategies that would allow a next-generation implant to decompose the visual scene into semantically meaningful parts. This has been done previously using depth (e.g., McCarthy et al., 2015), thermal sensors (e.g., Sadeghi et al., 2021), or computer vision (e.g., Jiang et al., 2020, Wang et al., 2014). However, this is the first study to use computer vision/semantic segmentation to break the scene into separate components and use time as a secondary avenue for separating the components into recognizable parts.

Taken together, this work has the potential to 1) further our understanding of the qualitative experience associated with different visual prosthetic technologies, 2) provide realistic expectations of prosthesis performance for patients, doctors, manufacturers, and regulatory bodies, and 3) accelerate the prototyping of new devices that may one day restore useful vision to people living with profound blindness.

1.3 Organization

The remainder of this thesis is organized as follows.

Chapter 2 provides some technical background on prosthetic vision, introducing different devices and describing the process by which they work. It discusses the benefits of simulation studies and facilitates a basic understanding for the field of SPV.

Chapter 3 introduces BionicVisionXR, the open-source toolbox used in Chapters 4–6 to study often neglected factors in SPV, the effects of different electrode rasterization patters, and the ability to use computer vision to not just spatially segment an image, but to temporally segment an information stream. The chapter gives an overview of BionicVisionXR's features, highlights architectural and algorithmic challenges overcome during development, analyzes performance, and discusses related work.

Chapter 4 summarizes a study performed with BionicVisionXR to investigate the effects of spatial distortions and VR HMDs on simulated task performance. It discusses the methodologies of two tasks, letter recognition and obstacle avoidance, along with results indicating that these previously neglected factors in SPV have a significant impact on subject performance.

Chapter 5 summarizes a study performed with BionicVisionXR to investigate the effects of different electrode rasterization patterns on task performance. This study introduces gaze congruent rendering and temporal effects in comparison to the previous study, but only briefly discusses these methodologies as they are primarily covered in Chapter 3. The methodologies and results are presented for two tasks, letter recognition and motion discrimination. The chapter ends by discussing biases imposed by certain rasterization patterns and that a "checkerboard" rasterization outperforms all other patterns of similar complexity in both subject performance and perceived difficulty.

Chapter 6 summarizes a study performed with BionicVisionXR to investigate the

effects of using computer vision to present different components of the scene at different times. It discusses the methodologies and results of an obstacle avoidance/wayfinding task in which subjects were required to identify their target location and avoid people, objects, and cyclists while navigating to their target. The chapter ends by discussing the performance benefits identified by the study, along with suggestions for future work to further utilize computer vision for prostheses users.

Chapter 7 presents the main conclusions defended in this thesis, discusses limitations of the work, and outlines possible avenues for future studies.

1.4 Permissions and Attributions

Chapter 2

Background

2.1 Visual Prosthetics

Few disabilities affect human life more than the loss of the ability to see. Although some affected individuals can be treated with surgery or medication, and recent advances in gene and stem cell therapies are showing great promise, there are no effective treatments for many people blinded by severe degeneration or damage to the retina, the optic nerve, or cortex. In such cases, an electronic visual prosthesis ("bionic eye", Figure 2.1) may be the only option (Fernandez, 2018).

Numerous groups worldwide are pursuing a visual prosthesis that stimulates viable neuronal tissue in the hope of generating functional vision artificially. Two devices have previously been approved for commercial use, but are no longer supported: Argus II (60 electrodes, Second Sight Medical Products, Inc., Luo and da Cruz, 2016a) and Alpha-IMS (1500 electrodes, Retina Implant AG, Stingl et al. 2013). In addition, PRIMA (378 electrodes, Pixium Vision, Lorach et al., 2015) has started clinical trials, with others to follow shortly (Ayton et al., 2014; Ferlauto et al., 2018). Table 2.1 shows all retinal and cortical implants that have made it to clinical trials, gotten FDA

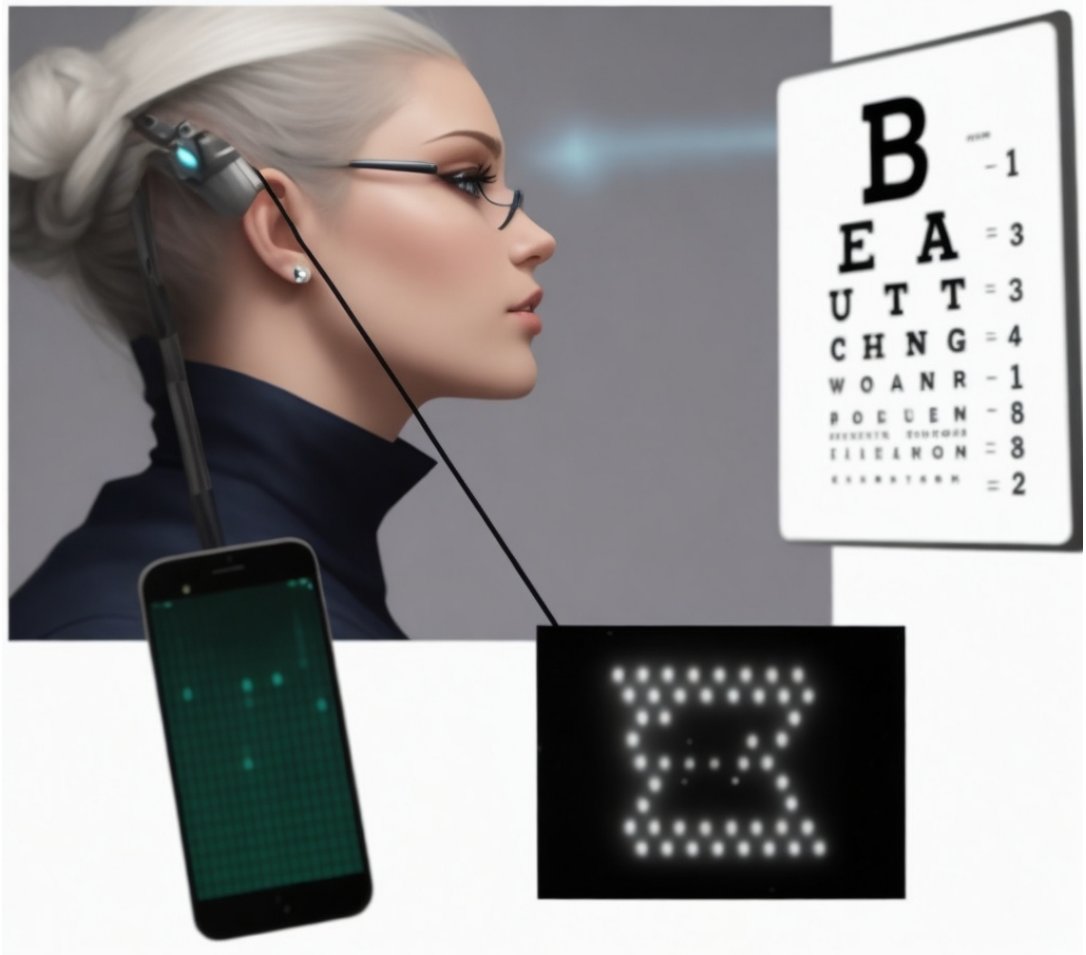


Figure 2.1: **Overview of bionic vision:** A camera, typically mounted to a pair of glasses, captures an image and sends it to the system's video processing unit. The image is processed and translated into a pattern of electrode activation values. These values are used to stimulate an implanted electrode array located in the brain, eye, or along the path from retina to cortex. This results in the user experiencing 'phosphenes', the perceptual experience of electrically stimulating visual neurons

approval, or CE mark approval. These devices acquire visual input via an external camera and perform simple image processing (e.g., contrast enhancement) via an external video processing unit (VPU), before sending the signal through wireless coils to a microstimulator implanted in the eye or the brain (see Figure 2.1). The stimulator receives the information, decodes it, and stimulates the visual system with electrical current, ideally creating artificial vision.

Implant	Status	Implant Location	Electrode Count	Number Implanted
Argus II	Inactive	Epiretinal	60	> 500
IRIS II	Inactive	Epiretinal	150	10
Polyretina	Pre-trial	Epiretinal	10,498	0
Alpha-AMS	Inactive	Subretinal	1600	15+
PRIMA	Clinical Trials	Subretinal	378	5+
Suprachoroidal 44	Clincal Trials	Suprachoroidal	44	4
Cortivis	Clinical Trials	Cortical	100	12+
ICVP	Clinical Trials	Cortical	16	5 (estimate)
Link	Trial Recruitment	Cortical	3072	0
Orion	Clinical Trials	Cortical	60	6
AV-Done	Clinical Trials	Optic Nerve	7	1

Table 2.1: For a complete list of devices in development, see `https://bionic-vision.org/devices`.

However, a major outstanding challenge in the use of these devices is translating electrode stimulation into a code that the brain can understand. The prosthetic vision generated by current retinal implants is still rudimentary and does not differ much across different device technologies (Erickson-Davis and Korzybska, 2021). Analogous to the first generation of cochlear implants, these devices have relied on straightforward signal processing and encoding schemes, assuming that each electrode in the array can be thought of as a "pixel" in an image (Dagnelie et al., 2007; Chen et al., 2009; Perez-Yus et al., 2017; Sanchez-Garcia et al., 2019); to generate a complex visual experience, one then simply needs to turn on the right combination of pixels.

In contrast, current prosthesis users report seeing highly distorted phosphenes, which vary in shape across subjects as well as electrodes and often fail to assemble into more complex percepts (Wilke et al., 2011; Beyeler et al., 2019b; Beauchamp et al., 2020; Erickson-Davis and Korzybska, 2021; Fernández et al., 2021). In the case of epiretinal implants, these distortions are largely due to inadvertent activation of passing axon fibers (Rizzo et al., 2003; Beyeler et al., 2019b), but other device technologies based on

electrical stimulation of visual cortex or optogenetics may face related issues. On the one hand, optogenetic prostheses may cause perceptual distortions due to differences in temporal dynamics between the optogenetic molecules and normal photopigments (Fine and Boynton, 2015). On the other hand, although there is a long history of patients reporting punctate percepts (sometimes described as "a star in the sky") in response to single-electrode stimulation of the visual cortex (Dobelle and Mladejovsky, 1974; Evans et al., 1979; Dobelle, 2000; Bosking et al., 2017), more recent work has highlighted that the percepts resulting from multi-electrode stimulation cannot be explained by a summative model based on single-electrode phosphenes (Beauchamp et al., 2020; Barry et al., 2020; Fernández et al., 2021).

Instead, a growing body of evidence suggests that individual electrodes do not lead to the perception of isolated, focal spots of light (Fine and Boynton, 2015; Beyeler et al., 2019b; Erickson-Davis and Korzybska, 2021). Although consistent over time, phosphenes vary drastically across subjects and electrodes (Luo et al., 2016; Beyeler et al., 2019b) and often fail to assemble into more complex percepts (Rizzo et al., 2003; Wilke et al., 2011). Consequently, retinal implant users do not see a perceptually intelligible world (Erickson-Davis and Korzybska, 2021).

2.2 Simulated Prosthetic Vision

However, a major outstanding challenge in the use of these devices is translating electrode stimulation into a code that the brain can understand. A common misconception is that each electrode in the grid can be thought of as a "pixel" in an image, and most existing devices linearly translate the grayscale value of a pixel in each video frame to a current amplitude of the corresponding electrode in the array. This is known as the scoreboard model, which implies that creating a complex visual scene can

be accomplished simply by using the right combination of pixels (analogous to creating numbers on a sports stadium scoreboard).

On the contrary, recent work suggests that phosphenes vary in shape and size, differing considerably across subjects and electrodes (Fine and Boynton, 2015; Luo et al., 2016; Beyeler et al., 2019b).

For instance, the perceived shape of phosphenes generated by an epiretinal implant was recently demonstrated to be a result of unintended stimulation of nerve fiber bundles (NFBs) in the retina (Rizzo et al., 2003; Beyeler et al., 2019b). These NFBs follow polar trajectories (Jansonius et al., 2012) away from the horizontal meridian, forming arch-like projections into the optic nerve (Fig. 2.2, *left*). Stimulating a NFB would result in the activation of nearby retinal ganglion cells (RGCs) that are upstream in the trajectory, resulting in phosphenes that appear elongated (Fig. 2.2, *right*).

Beyeler et al. (2019b) demonstrated through simulations that the shape of elicited

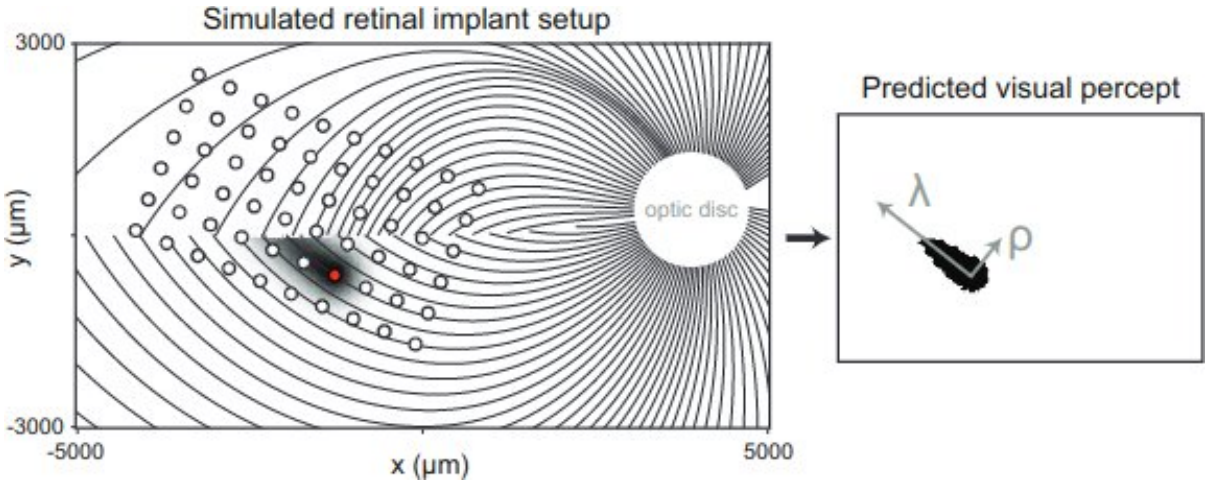


Figure 2.2: A simulated map of retinal NFBs (*left*) can account for visual percepts (*right*) elicited by retinal implants (reprinted with permission from Beyeler et al. (2019a)). ***Left*:** Electrical stimulation (red circle) of a NFB (black lines) could activate retinal ganglion cell bodies peripheral to the point of stimulation, leading to tissue activation (black shaded region) elongated along the NFB trajectory away from the optic disc (white circle). ***Right*:** The resulting visual percept appears elongated; its shape can be described by two parameters, λ and ρ.

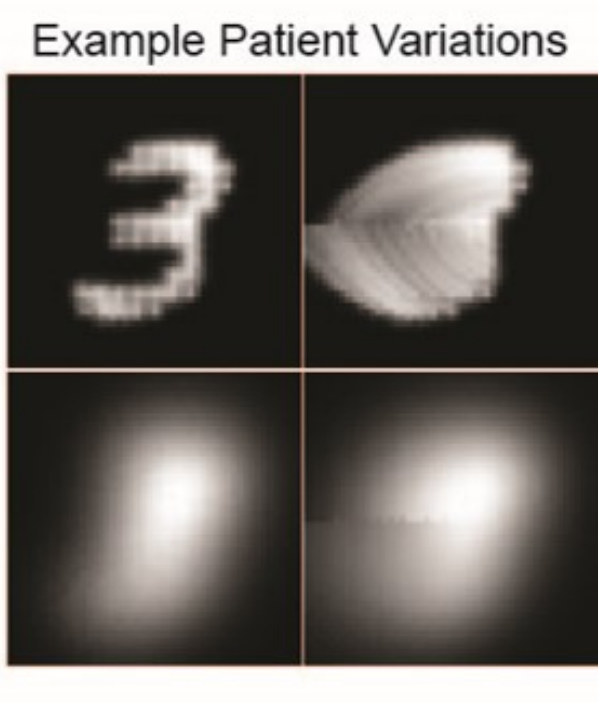


Figure 2.3: The number '3' with increasing values of rho (ρ) and lambda (λ). Higher values increase distortion radially (ρ) and along the axonal trajectories (λ).

phosphenes closely followed NFB trajectories. Their computational model assumed that an axon's sensitivity to electrical stimulation:

i. decayed exponentially with ρ as a function of distance from the stimulation site,

ii. decayed exponentially with λ as a function of distance from the cell body, measured as axon path length.

In other words, the values of ρ and λ in this model dictate the size and elongation of phosphenes, respectively. This may drastically affect visual outcomes, as large values of λ are thought to distort phosphene shape (see Figure 2.3).

2.3 Virtual Reality

Despite their potential to restore vision to people living with incurable blindness, the number of bionic eye users in the world is still relatively small ($\sim$ 500 retinal prostheses implanted to date). To investigate functional recovery and experiment with different implant designs, researchers have therefore been developing VR prototypes that rely on SPV.

The classical method relies on sighted subjects wearing a VR headset, who are then deprived of natural viewing and only perceive phosphenes displayed in a HMD. This viewing mode has been termed *transformative reality* Lui et al. (2011) (as opposed to *altered reality* typically used to describe low vision simulations Bao and Engel (2019)), which allows sighted users to "see" through the eyes of the bionic eye recipient, taking into account their head and/or eye movements as they explore a virtual environment Kasowski et al. (2021).

A number of previous VR studies have focused on assessing the impact of different stimulus and model parameters (e.g., phosphene size, phosphene spacing, flicker rate) on measures of visual acuity. Stimuli for these low-level visual function tests were often presented on monitors (Vurro et al., 2014; Lu et al., 2012) or in HMDs (Wu et al., 2014; Cao et al., 2017; Caspi and Zivotofsky, 2015). Some studies also tested the influence of FOV (Sanchez-Garcia et al., 2020; Thorn et al., 2020) and eye gaze compensation (Titchener et al., 2018) on acuity. Others focused on slightly more complex tasks such as letter (Zhao et al., 2011), word (Fornos et al., 2011), face (Denis et al., 2013; Chang et al., 2012), and object recognition (Zhao et al., 2010; Wang et al., 2018; Macé et al., 2015). In most setups, participants would view SPV stimuli in a conventional VR HMD, but some studies also relied on smart glasses to present SPV in augmented reality (AR).

However, because most SPV studies rely on the scoreboard model (Dagnelie et al., 2007; Chen et al., 2009; Lui et al., 2011; Perez-Yus et al., 2017; Sanchez-Garcia et al., 2019), it is unclear how their findings would translate to real bionic eye recipients. Only a handful of studies have incorporated a great amount of neurophysiological detail into their setup (Josh et al., 2013; Vurro et al., 2014; Wang et al., 2018; Thorn et al., 2020; Kasowski and Beyeler, 2022), only three of which relied on an established and psychophysically validated model of SPV (Wang et al., 2018; Thorn et al., 2020; Kasowski and Beyeler, 2022). In addition, being able to move around as one would in real life has shown to significantly increase the level of immersion a user experiences (Pasch et al., 2009). However, the level of immersion offered by most SPV studies is relatively low, as stimuli are often presented on a screen (Ying et al., 2018; Wang et al., 2018). In contrast, most current prostheses provide a very limited FOV (e.g., Argus II: 10×20 degrees of visual angle), which requires users to scan the environment with strategic head movements while trying to piece together the information (Erickson-Davis and Korzybska, 2021). Furthermore, Argus II does not take into account the eye movements of the user when updating the visual scene, which can be disorienting for the user. Ignoring these human-computer interaction (HCI) aspects of bionic vision can result in unrealistic predictions of prosthetic performance. It is therefore unclear how the findings of most SPV studies would translate to real prosthesis users.

Chapter 3

BionicVisionXR: An Open-Source Virtual Reality Toolbox for Bionic Vision

3.1 Introduction

Due to the unique requirements of working with bionic eye recipients (e.g., constant assistance, increased setup time, travel cost), experimentation with different encoding methods remains challenging and expensive. Instead, embedding models of SPV in immersive VR allows sighted subjects to act as *virtual patients* by "seeing" through the eyes of the patient, taking into account their head and eye movements as they explore an immersive virtual environment. This can speed up the development process by allowing researchers to test theoretical predictions in high-throughput experiments, the best of which can be validated and improved upon iteratively with the bionic eye recipient in the loop. However, very few studies attempted to model bionic vision in real time or use spatial distortions found in real prosthesis users (see Section 2.3).

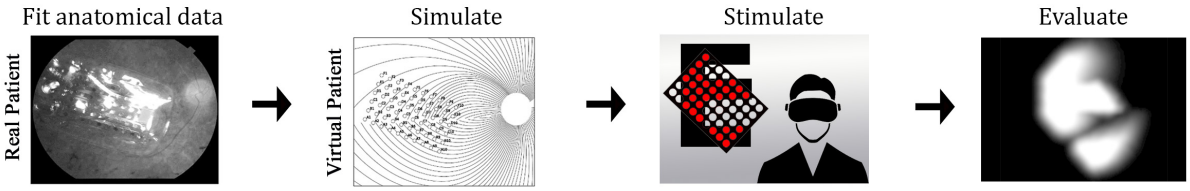


Figure 3.1: **Simplified overview of the simulated prosthetic vision model:** Anatomical data from a real patient is used to simulate perception. A virtual environment is used to present stimulus to the model, which renders simulated prosthetic vision to the participant. The participant is treated as a "virtual patient", and their performance on common tasks is evaluated.

3.2 BionicVisionXR

To address these challenges, I developed BionicVisionXR, an open-source virtual reality toolbox for simulated prosthetic vision that uses a neurophysiologically inspired and psychophysically validated computational model to accurately model the predicted experience of a prosthesis user (see Figure 3.1). The entire software package is a combination of C#, HLSL (shader language), and calls to the python package *pulse2percept* (Beyeler et al., 2019b). The project is open source and available at `https://github.com/bionicvisionlab/BionicVisionXR`.

The general workflow is as follows:

1) Image acquisition: Unity's virtual camera captures a 60-degree field of view at 90 frames per second.

2) Image processing: The image is typically downscaled, converted to greyscale, preprocessed, and blurred with a Gaussian kernel. Preprocessing includes things like depth or edge detection, contrast enhancement, etc.

3) Electrode activation: Electrode activation is derived directly from the closest pixel to each electrode's location in the visual field. The previous blurring is in place to avoid misrepresenting crisp edges, where moving one pixel could result in an entirely different activation value. Activation values are only collected for electrodes that are

currently active (see Section 5.1.3).

4) Spatial effects: The electrode activation values are used with a psychophysically validated phosphene model (Beyeler et al., 2019b) to determine the brightness value for each pixel in the current frame (see Section 3.2.1).

5) Temporal effects: Previous work has demonstrated phosphene fading (Fornos et al., 2012) and persistence (Schmidt et al., 1996) for epiretinal devices. Additionally, previous simulations have eluded to the importance of temporal properties in electrode stimulation strategies (Avraham et al., 2021; Thorn et al., 2022). To simulate these effects, we implemented a charge accumulation and decay model with parameters matching previously reported temporal properties in real devices. Information from previous frames is used to adjust the brightness of subsequent frames (see Section 3.2.2).

6) Gaze-contingent rendering: Gaze contingency (when what you're being shown is congruent with your gaze) significantly improves performance on various tasks using real devices (Caspi et al., 2018) and simulated prosthetic vision (Titchener et al., 2018; Bourkiza et al., 2013; Rassia and Pezaris, 2018; Paraskevoudi and Pezaris, 2019). The package has the option to access device eye-trackers and present the stimulus as either gaze congruent or incongruent (see Section 3.2.3).

Each of these steps is described in more detail below:

3.2.1 Spatial Distortions

The shape of phosphenes in epiretinal devices is thought to be influenced by the underlying neuroanatomy of the retina. These devices aim to stimulate retinal ganglion cells on the epiretinal surface. However, as mentioned in Section 2.2, these cells send axonal projections to the optic nerve in arch-like trajectories. As the device stimulates

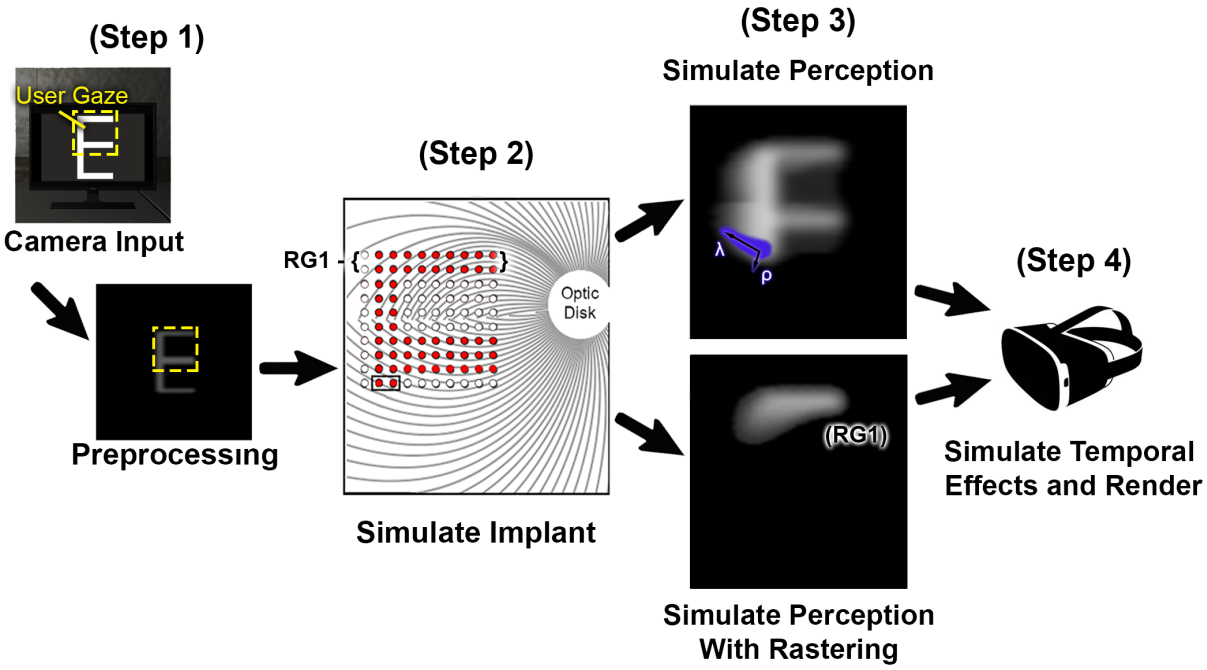


Figure 3.2: **Simplified overview of the simulated prosthetic vision model: Step 1 -** Unity's camera view of the virtual scene is captured, preprocessed, and used to simulate neuronal stimulation. **Step 2 -** The simulated device is centered over the user's gaze and neuronal activation is simulated. The neuroanatomy of the retina is used to predict phosphenes elongated along simulated axonal trajectories. Raster group 1 (RG1) electrodes are shown in brackets. **Step 3 -** Perception is simulated without rastering (top), or with electrode rastering (bottom). Two terms, ρ (representing phosphene spread) and λ (representing phosphene elongation) are used to simulate phosphene shape. The highlighted shape represents the predicted percept of the two active electrodes in the bottom row. In the rastered condition, only 20/100 electrodes are active on any given frame. In this example, only the top two rows (Raster Group 1) are active. **Step 4 -**Temporal dynamics are modeled and the final image is rendered to the headset.

nearby retinal ganglion cells, it inadvertently stimulates the axons of retinal ganglion cells that traverse the same space (Fig. 2.2), which may result in phosphenes elongated along the axonal arches (Beyeler et al., 2019b).

To model phosphene shape, we used the psychophysically validated axon map model (Beyeler et al., 2019b), which assumed that an axon's sensitivity to electrical stimulation decays exponentially as a function of (i) distance from the stimulating electrode (with decay rate ρ), and (ii) distance along the axon from the cell body (with

decay rate λ). The output of the model was an intensity profile $b(r, \theta)$, representing the perceived brightness of a phosphene at each pixel ((r, θ) describing the corresponding polar coordinates in the retina, centered over the fovea):

$$b(r, \theta) = max_{p \in R(\theta)} \sum_{e \in E} \exp\left(\frac{-d_e^2}{2\rho^2} + \frac{-d_{\text{soma}}^2}{2\lambda^2}\right), \tag{3.1}$$

where $R(\theta)$ was the path of the axon to the point (r, θ), p was an individual point along the path, d_e was the Euclidean distance from p to the stimulating electrode, E was the set of all electrodes, and d_{soma} was the distance from p to the cell body along the axon, given by the path integral over the nerve fiber bundle:

$$d_{\text{soma}} = \int_p^\theta \sqrt{R(\theta)^2 + \left(\frac{dR(\theta)}{d\theta)}\right)^2} d\theta. \tag{3.2}$$

The paths of these nerve fiber bundles ($R(\theta)$) were modeled as spirals originating the optic disc and terminating at each ganglion cell body (Beyeler et al., 2019b).

3.2.2 Temporal Distortions

In an effort to capture the temporal dynamics, we employed a simpler, linear variant of equations drawn from an established model (Horsager et al., 2009). The model contains two leaky integrators representing neuron desensitization, $n(t) \in [0, \infty)$, and phosphene brightness, $b(t) \in [0, \infty)$, respectively, at each pixel location:

$$\frac{dn(t)}{dt} = -\tau_n n(t) + b_I(t), \tag{3.3}$$

$$\frac{db(t)}{dt} = -\tau_b b(t) - \alpha n(t) + b_I(t) \tag{3.4}$$

The model included brightness input ($b_I(t)$, previously calculated as $b(r, \theta)$ as described

in Eq. 3.1), desensitization decay (τ_n), brightness decay (τ_b), and a desensitization scaling factor (α). The temporal model was performed on every pixel of the downscaled image. The values for τ_b, τ_n, and α can be modified to fit patient data based on the stimulation parameters being used.

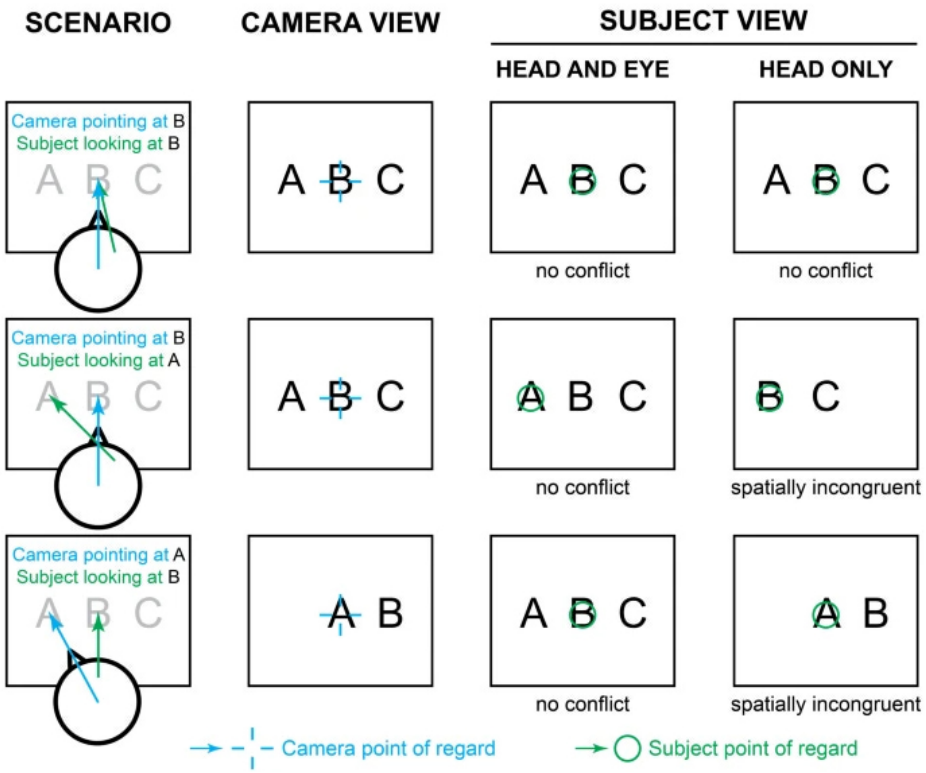


Figure 3.3: A demonstration of gaze congruency with a head mounted camera. If the camera is aligned with the user's gaze, there is no conflict (top row). If the camera is pointed at B, but the user is gazing to their left, at the letter A, they will perceive the letter B as being shifted to the left (middle row). Similarly, if the camera is pointed to the left at the letter A, but the user is gazing at the letter B, they will perceive the letter A at the location of the letter B (bottom row). Reused under CC-BY from Paraskevoudi and Pezaris (2021).

3.2.3 Gaze-Contingent Phosphene Rendering

Spatial updating is a process that relies on information from eye movements to understand the spatial layout of a visual scene. This process depends on something

known as gaze congruency (see Fig 3.3), an especially important factor for prosthetic vision (see Paraskevoudi and Pezaris (2019) for a detailed review). Gaze congruency has been shown to significantly affect task performance both in physical epiretinal devices (Sabbah et al. (2014); Barry and Dagnelie (2016); Caspi et al. (2021)), and simulation studies (Bourkiza et al. (2013); Rassia and Pezaris (2018); Paraskevoudi and Pezaris (2021))

To implement gaze congruency, the image being processed underwent two shifts (see Figure 3.4). First, the input image was shifted in the opposite direction of the user's gaze (in screen coordinates), effectively moving the user's fixation point to the center for processing. This shift was performed to center the image over the simulated electrode array (centered at $x = 0$, $y = 0$), allowing for pre-computed electrode-pixel interactions to be used and insuring the temporal effects were applied to retinal locations and not screen coordinates. Electrode activation patterns were extracted from the shifted pixel intensities and used in the previously described methods. After modeling the temporal effects, a final gaze shift was performed to reposition the simulated percepts to the user's focus point in real time. As a result, the participant perceived a moving "window" of phosphenes that shifted with their gaze.

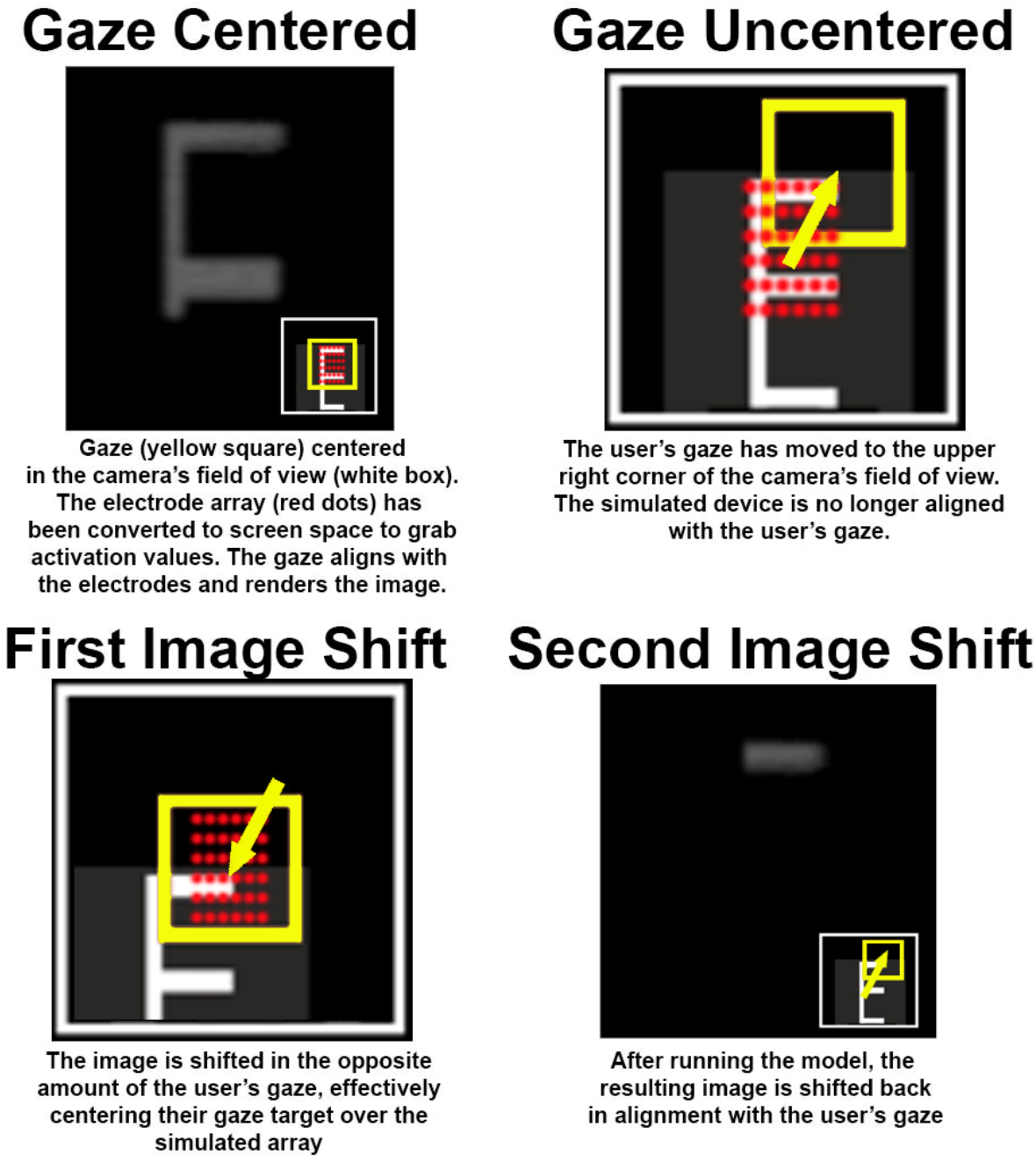

Gaze Centered
Gaze (yellow square) centered in the camera's field of view (white box). The electrode array (red dots) has been converted to screen space to grab activation values. The gaze aligns with the electrodes and renders the image.
Gaze Uncentered
The user's gaze has moved to the upper right corner of the camera's field of view. The simulated device is no longer aligned with the user's gaze.
First Image Shift
The image is shifted in the opposite amount of the user's gaze, effectively centering their gaze target over the simulated array
Second Image Shift
After running the model, the resulting image is shifted back in alignment with the user's gaze

Figure 3.4:

Chapter 4

Immersive Virtual Reality Simulations of Visual Prostheses

4.1 Introduction

While many simulation studies have focused on prosthetic vision, few results have been replicated with real device users. In many cases, simulation studies found performance on certain tasks far exceeded the performance when the task was performed by prosthesis users. Conversely, some prosthesis users were able to surpass the theoretical acuity limits of their device. There are many potential causes for these discrepancies, but two likely contributions are inaccurate phosphene shapes and the unrealistic presentation of prosthetic vision on a computer monitor (see Section 2.2).
We therefore systematically evaluated how different display types (HMD or monitor) affect behavioral performance in a letter recognition and an obstacle avoidance task. To the best of our knowledge, this is the first SPV study that uses a within-subjects design to allow for a direct comparison between display types of the same tasks.

4.2 Methods

4.2.1 Participants

We recruited 17 sighted participants (6 female and 11 male; ages 27.4 ± 5.7 years) from the student pool at the University of California: Santa Barbara. Participation was voluntary and subjects were informed of their right to freely withdraw for any reason. Recruitment and experimentation followed protocols approved by the university's Institutional Review Board, along with limitations and safety protocols approved by the university's COVID-19 response committee.

4.2.2 Simulated Prosthetic Vision

BionicVisionXR was used to render phosphenes in real time. The shape of the elicited phosphenes was based on the retinal location of the simulated implant as well as model parameters ρ and λ (see Section 3.2.1). As can be seen in Fig. 2.2 (*left*), electrodes near the horizontal meridian activated cells close to the end of the NFBs, limiting the potential of elongation along an axon. This resulted in more circular phosphenes, whereas other electrodes were predicted to produce elongated percepts that differed in angle based on whether they fell above or below the horizontal meridian.
We were particularly interested in assessing how different SPV model parameters affected behavioral performance. Importantly, ρ and λ vary drastically across patients Beyeler et al. (2019b). Although the reason for this is not fully understood, it is clear that the choice of these parameter values may drastically affect the quality of the generated visual experience. To cover a broad range of potential visual outcomes, we simulated nine different conditions by combining $\rho = \{100, 300, 500\}$ µm with $\lambda = \{50, 1000, 5000\}$µm.

We were also interested in how the number of electrodes in an implant and the associated change in FOV affected behavioral performance. In addition to simulating Argus II, we created two hypothetical near-future devices that used the same aspect ratio and electrode spacing, but featured a much larger number of electrodes. Thus the three devices tested were:

- Argus II: $6 \times 10 = 60$ equally spaced electrodes situated 575 µm apart in a rectangular grid. To match the implantation strategy of Argus II, the device was simulated at $-45°$ with respect to the horizontal meridian in the dominant eye.
- Argus III (hypothetical): $10 \times 16 = 160$ electrodes spaced 575 µm apart in a rectangular grid implanted at 0°. A recent modeling study suggests that this implantation angle might minimize phosphene streaks Beyeler et al. (2019a).
- Argus IV (hypothetical): $19 \times 31 = 589$ electrodes spaced 575 µm apart in a rectangular grid implanted at 0°.

4.2.3 Tasks

To study the impact of SPV parameters and level of immersion, we replicated two popular tasks from the bionic vision literature. The first task was a basic letter recognition experiment (Cruz et al., 2013), tasking participants with identifying the letter presented to them. The second one was a more immersive orientation & mobility task, requiring subjects to walk down a virtual hallway while avoiding obstacles (He et al., 2020).

To allow for a direct comparison across all conditions, we chose a within-subjects, randomized block design. This systematic side-by-side comparison minimized the risk of learning effects and other artifacts that may arise from inhomogeneity between groups, allowing for meaningful statistics with a relatively small number of subjects.

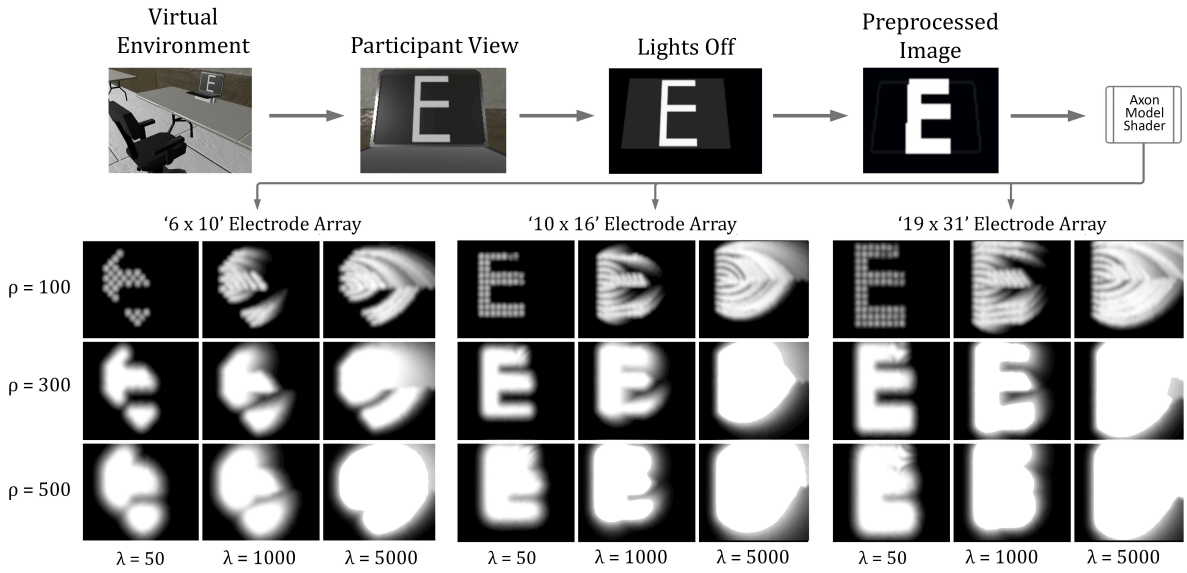


Figure 4.1: Letter recognition task. ***Top:*** The lights in the virtual room are turned off and the image seen by the user is passed to the preprocessing shader which performs edge extraction/enhancement before the axon model shader renders SPV. Modeled after Cruz et al. (2013). ***Bottom:*** Output of the axon model shader across the various devices and ρ / λ combinations.

The procedures and results for each task are presented separately below, followed by a joint discussion on both experiments in the subsequent sections.

Task 1: Letter Recognition

Original Task The first experiment was modeled after a letter recognition task performed by Argus II recipients (Cruz et al., 2013). In the original task, following a short training period, participants were instructed to identify large and bright white letters presented on a black screen situated 0.3 m in front of them. Participants were given unlimited time to respond. The experiment was carried out in a darkened room. Both the initial training period and the actual experiment featured all 26 letters of the alphabet. The letters were grouped by similarity and tested in batches of 8, 8, and 10 letters.

Experimental Setup and Procedure To emulate the experiment described in Cruz et al. (2013), we carefully matched our virtual environment to the experimental setup of the original task. The setup mainly consisted of a virtual laptop on top of a virtual desk (Fig. 4.1). A virtual monitor was positioned 0.3 m in front of the user's head position. In agreement with the original task, participants were presented letters that were 22.5 cm tall (subtending 41.112° of visual angle) in True Type Century Gothic font. For the monitor version of the task, the camera was positioned at the origin and participants could simulate head movements by using the mouse.
Each combination of 3 devices $\times$ 3 ρ values $\times$ 3 λ values were implemented as a block, resulting in a total of 27 blocks. All 27 blocks were completed twice; once for the HMD version of the task, and once for the monitor version of the task. Rather than presenting all 26 letters of the alphabet (as in the original experiment), we limited our stimuli to the original Snellen letters (C, D, E, F, L, O, P, T, Z) for the sake of feasibility.
All nine Snellen letters were presented in each block, resulting in a total of 243 trials. Participants were limited to 1 minute per trial, after which the virtual monitor would go dark and the participant had to select a letter before the experiment continued.
To acclimate participants to the task and controls, we had them perform an initial practice trial using normal vision. After that, the lights in the virtual room were turned off and BionicVisionXR was used to generate SPV. To mimic the training session of Cruz et al. (2013), participants completed three practice trials using SPV at the beginning of each block. Participants were able to repeat each practice trial until they had selected the correct letter. To prevent participants from memorizing letters seen during practice trials, we limited practice trials to the letters Q, I, and N.
Participant responses and time per trial were recorded for the entirety of the experiment.

Performance Evaluation Perceptual performance was assessed using F1 scores, which represent the harmonic mean between precision and recall, allowing for a slight penalty towards false positive choices compared to recall (proportion correct) on its own. This had the advantage of eliminating bias towards specific letter choices. F1 values were calculated for each block using the scikit-learn 'f1_score' function (Pedregosa et al., 2011). We also measured time per trial with the assumption that easier trials could be completed faster than trials that were more difficult.
Due to ceiling and floor effects, neither outcome measure (F1 scores and time per trial) were normally distributed, violating the assumptions of the standard ANOVA. We therefore performed a subsequent aligned rank transform (ART) with the R package ARTool for both F1 scores and time per trial. This method of analysis allows for a factorial ANOVA to be performed on repeated measures, non-uniform data, and lower subject counts (Wobbrock et al., 2011). Post-hoc analyses were performed on significant groups by analyzing the rank-transformed contrasts (Elkin et al., 2021). The Tukey method (Tukey, 1949) was used to adjust p-values to correct for multiple comparisons. All code used in the analysis, along with the raw data, is provided at `https://github.com/bionicvisionlab/2022-kasowski-immersive`.

Task 2: Obstacle Avoidance

Original Task The second task was modeled after an obstacle avoidance experiment performed by Argus II recipients (He et al., 2020). In this task, participants were required to walk down a crowded hallway with one to three people located at one of four fixed distances on either the left or right side of the hallway. Participants were permitted the use of a cane and were allowed to touch the walls with the cane (but not the standing persons). Participants were given unlimited time to complete the task and were closely monitored by the experimenter to avoid dangerous collisions. For each

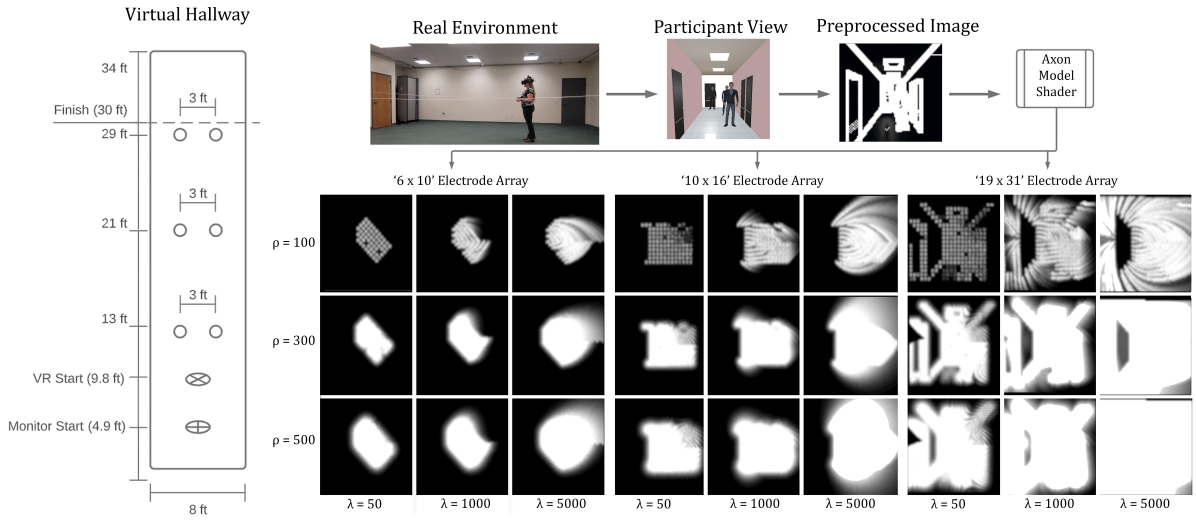


Figure 4.2: Obstacle avoidance task. ***Left:*** Layout of the virtual hallway environment modeled after He et al. (2020). Empty circles represent the possible locations for obstacles. ***Right/Top:*** View of the real environment -¿ participant's view is passed to the preprocessing shader which performs edge extraction/enhancement before the axon model shader renders SPV. ***Bottom:*** Output of the axon model shader across the various devices and ρ / λ combinations.

trial, the experimenter instructed the participant to stop when they reached the end of the hallway.

Experimental Setup and Procedure To emulate the experiment described in He et al. (2020), we designed a virtual hallway (Fig. 4.2, *Left*) modeled closely after the description and pictures of the physical hallway.

Participants were tasked with successfully navigating the virtual hallway while avoiding collisions with obstacles (simulated people). Each trial consisted of navigating past either two or three obstacles (three trials per condition, six trials total) located on either the left or right side of the hallway (Fig. 4.2).

To acclimate participants to the task and controls, we had them perform three initial practice rounds using normal vision. After that, participants completed three more practice rounds with a high-resolution scoreboard model (31×19 electrodes,

$\rho = 100\,\mu m$, $\lambda = 50\,\mu m$). Participants were instructed to complete the trials as quickly as possible while avoiding collisions. They were informed that collisions would result in audio feedback; a sample of each sound was played at the beginning of the experiment. Each combination of 3 devices $\times$ 3 ρ values $\times$ 3 λ values were implemented as a block, resulting in a total of 27 blocks. Block order was randomized and participants completed six trials per block for a total of 162 trials for each version (HMD/monitor) of the task. Participants were limited to 1 minute per trial, after which vision was returned to normal and participants walked to the end of the hallway to begin the next trial.

To ensure the safety of participants during the HMD-based version of the task, we positioned rope at the real-life location corresponding to each wall of the hallway (Fig 4.2, *Top, Left*). The rope served to guide the participants safely along the path while keeping them in bounds, but was also a substitution for the cane usage in the previous research. This substitution was necessary, because our testing facility was much larger than the hallway in the original experiment; thus the virtual walls did not coincide with physical walls.

An experimenter was always nearby to ensure the safety of the participants but did not otherwise interact with them during the experiment. At the end of each trial, the screen turned red and on-screen text instructed participants to turn around and begin the next trial in the other direction.

The monitor version of the task was similar, but each new trial would start automatically without the subject needing to turn around. Participants were seated in front of a monitor and were able to use the keyboard to move and the mouse to look around. The size of the hallway and positions of the obstacles were identical between versions, but participants started 1.5m closer to the first obstacle in the HMD version due to size restrictions of the room.

Collisions were detected using Unity's standard continuous collision detection software, with each obstacle having a $0.7\,\text{m} \times 0.4\,\text{m}$ hitbox and the participant having a radius of $0.4\,\text{m}$. Subject locations and orientations were continuously recorded. Time per trial, along with individual positions and timings of each collision, were recorded for each trial.

Evaluating Performance Performance was assessed by counting the number of collisions per trial and the amount of time to complete a trial, with a lower number of collisions or lower time per trial expected on easier trials. Analogous to the first task, these two metrics were averaged across trials in a block for each subject and analyzed using ART ANOVA. Post-hoc analyses were performed on significant groups using the Tukey method for multiple comparison adjustments.

4.2.4 Procedure

None of the participants had previous experience with SPV. Participants were split into two equally sized groups; one starting with the HMD-based version of the first experiment while the other started with the monitor-based version.
In order to get accommodated with the SPV setup, participants began each task with the easiest block; that is, the scoreboard model (λ=$50\,\mu\text{m}$) with the smallest possible phosphene size and the highest number of electrodes. The order of all subsequent blocks was randomized for each participant.

4.3 Results

4.3.1 Task 1: Letter Recognition

Results from the letter recognition task are summarized in Table 4.1 and distributions are plotted in Fig. 4.3. Group F-values, along with their significance, are reported in Table 4.2. Each data point in Fig. 4.3 represents a subject's F1 score (Fig. 4.3A–C) and time per trial (Fig. 4.3D–F) across all letters in a block. F1 score ranged from 0 to 1 with higher values representing better performance. Assuming a different letter is chosen for each selection, a chance-level F1 score would equal the probability for randomly guessing the correct letter ($\frac{1}{9} = 0.1111$).

	F1 Score (± Std Dev)		Mean Time (s) (± Std Dev)	
	HMD	Monitor	HMD	Monitor
06x10 Array	0.411 (± 0.341)	0.344 (± 0.339)	8.312 (± 5.685)	7.979 (± 7.367)
10x16 Array	0.628 (± 0.361)	0.546 (± 0.380)	5.853 (± 4.808)	**5.567 (± 3.608)**
19x31 Array	**0.699 (± 0.347)**	**0.596 (± 0.373)**	**5.379 (± 4.164)**	5.661 (± 4.373)
ρ=100	0.570 (± 0.388)	0.467 (± 0.381)	7.415 (± 5.874)	7.007 (± 6.120)
ρ=300	**0.620 (± 0.366)**	**0.540 (± 0.379)**	6.173 (± 4.879)	**5.829 (± 4.686)**
ρ=500	0.548 (± 0.352)	0.479 (± 0.377)	**5.956 (± 4.267)**	6.371 (± 5.483)
λ=50	**0.824 (± 0.267)**	**0.750 (± 0.329)**	**4.540 (± 3.408)**	**5.034 (± 4.403)**
λ=1000	0.665 (± 0.331)	0.543 (± 0.362)	5.698 (± 4.338)	6.074 (± 5.780)
λ=5000	0.248 (± 0.229)	0.193 (± 0.188)	9.307 (± 5.906)	8.099 (± 5.700)

Table 4.1: Letter recognition task: Average performance and time per trial across conditions. Best performances (highest F1/shortest time) for each grouping are presented in bold.

As expected, increasing the number of electrodes significantly increased F1 scores in both HMD (light gray) and monitor (dark gray) versions of the task (Fig. 4.3). It is worth noting that participants were consistently above chance levels, even with the simulated Argus II (6×10 electrodes) device. Increasing the number of electrodes also decreased the time it took participants to identify the letter (Fig. 4.3D). However, increasing the number of electrodes from 10×16 to 19×31 did not further decrease recognition time.

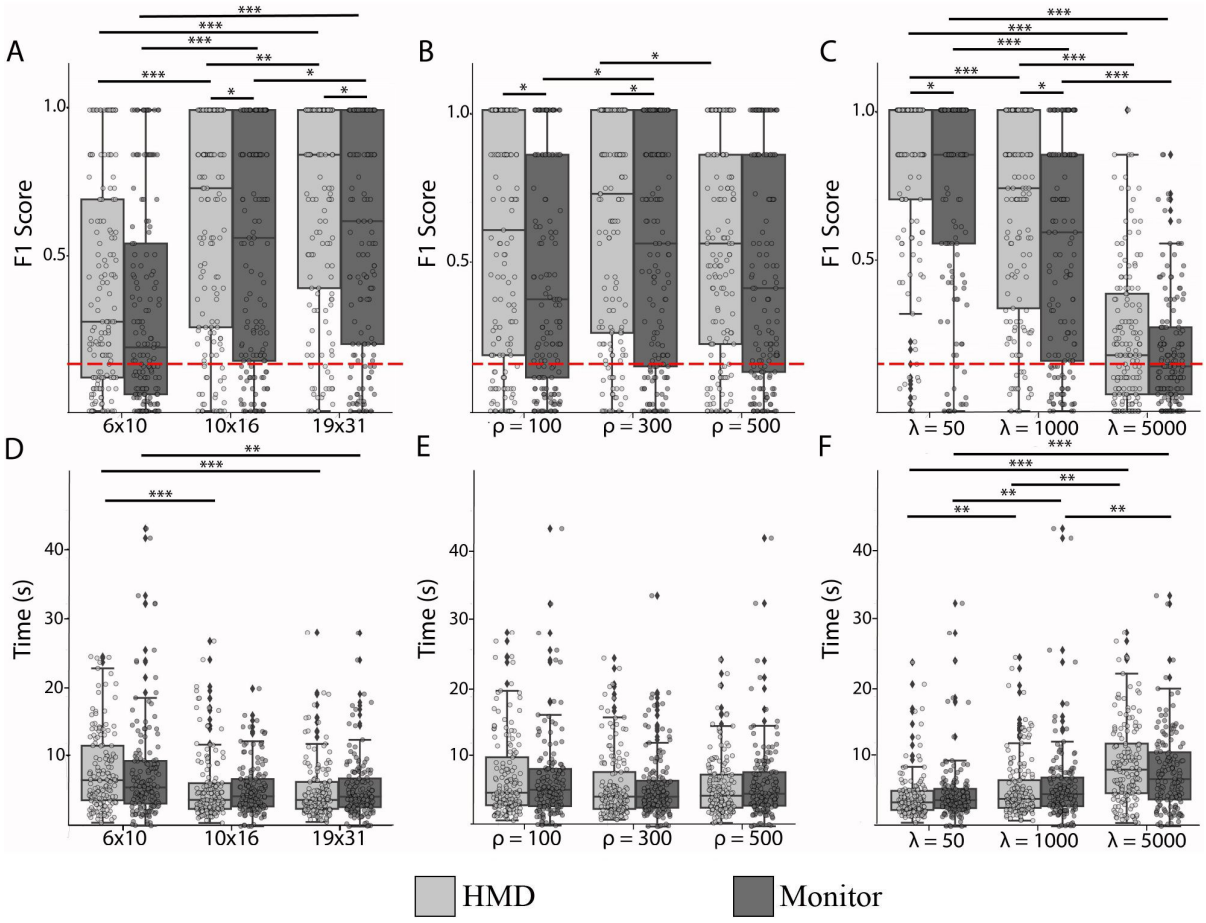


Figure 4.3: Letter recognition task. Data points represent each subject's average performance in a block with boxplots displaying median and interquartile ranges. ***Top:*** Average F1 score across blocks for each subject within the condition specified by the x-axis. ***Bottom***: Average time across blocks for each subject within the condition specified by the x-axis. Statistical significance was determined using ART ANOVA (*< .05, **< .01, ***< .001).

Contrary to previous findings, F1 scores and recognition time did not systematically vary as a function of phosphene size (ρ, Fig. 4.3B, E). In both HMD and monitor-based conditions, median F1 scores were highest for $\rho = 300\,\mu m$ (Table 4.1). However, participants achieved similar scores with $\rho = 100\,\mu m$ in the HMD version and with $\rho = 500\,\mu m$ in the monitor-based version of the task.

The most apparent differences in performance were found as a function of phosphene

elongation (λ, Fig. 4.3C, F). Using $\lambda = 50\,\mu m$, participants achieved a perfect median F1 score of 1.0, but this score dropped to 0.741 for $\lambda = 1000\,\mu m$ and 0.185 for $\lambda = 5000\,\mu m$ (Table 4.1). Increasing λ also significantly increased the time it took participants to identify the letter.

	F1 Score		**Time**	
	F-Value	Signif.	F-Value	Signif.
device	150.8174	9.17E-57	25.1232	2.51E-11
ρ	9.9004	5.62E-05	8.6049	2.00E-04
λ	535.8116	3.60E-151	80.6779	8.42E-33
display	31.5610	2.62E-08	1.4799	2.24E-01
device : ρ	0.7838	5.36E-01	0.5371	7.09E-01
device : λ	18.2971	1.98E-14	5.6673	1.67E-04
ρ : λ	10.0737	5.72E-08	0.5573	6.94E-01
device : display	0.3742	6.88E-01	1.3682	2.55E-01
ρ : display	0.2668	7.66E-01	0.5586	5.72E-01
λ : display	1.9499	1.43E-01	5.1031	6.27E-03
device : ρ : λ	1.3828	2.00E-01	2.4198	1.38E-02
device : ρ : display	0.3410	8.50E-01	0.3107	8.71E-01
device : λ : display	0.3956	8.12E-01	0.1496	9.63E-01
ρ : λ : display	0.7717	5.44E-01	0.6527	6.25E-01
device : ρ : λ : disp	0.4598	8.84E-01	0.6592	7.28E-01

Table 4.2: Letter recognition task: F-value table for Aligned Rank Transform (ART) ANOVA. Values were calculated with the ARTool software package. "device" refers to the three simulated electrode grids, while "display" refers to the use of an HMD or monitor.

A trend toward a higher F1 score when using the HMD was observed across all conditions (Fig. 4.3, *Top*), but the trend failed to reach significance for the device with the lowest number of electrodes (6×10 array) or across the larger distortion parameters (ρ=1000 µm and λ=5000 µm) (Fig. 4.3, *Top*). While average time per trial was faster across all conditions with the HMD, the effect was not significant (Fig. 4.3, *Bottom*).

4.3.2 Task 2: Obstacle Avoidance

Results are summarized in Table 4.3 and Fig. 4.4. Each data point in Fig. 4.4 represents a subject's number of collisions (Fig. 4.4, *Top*) and time to completion

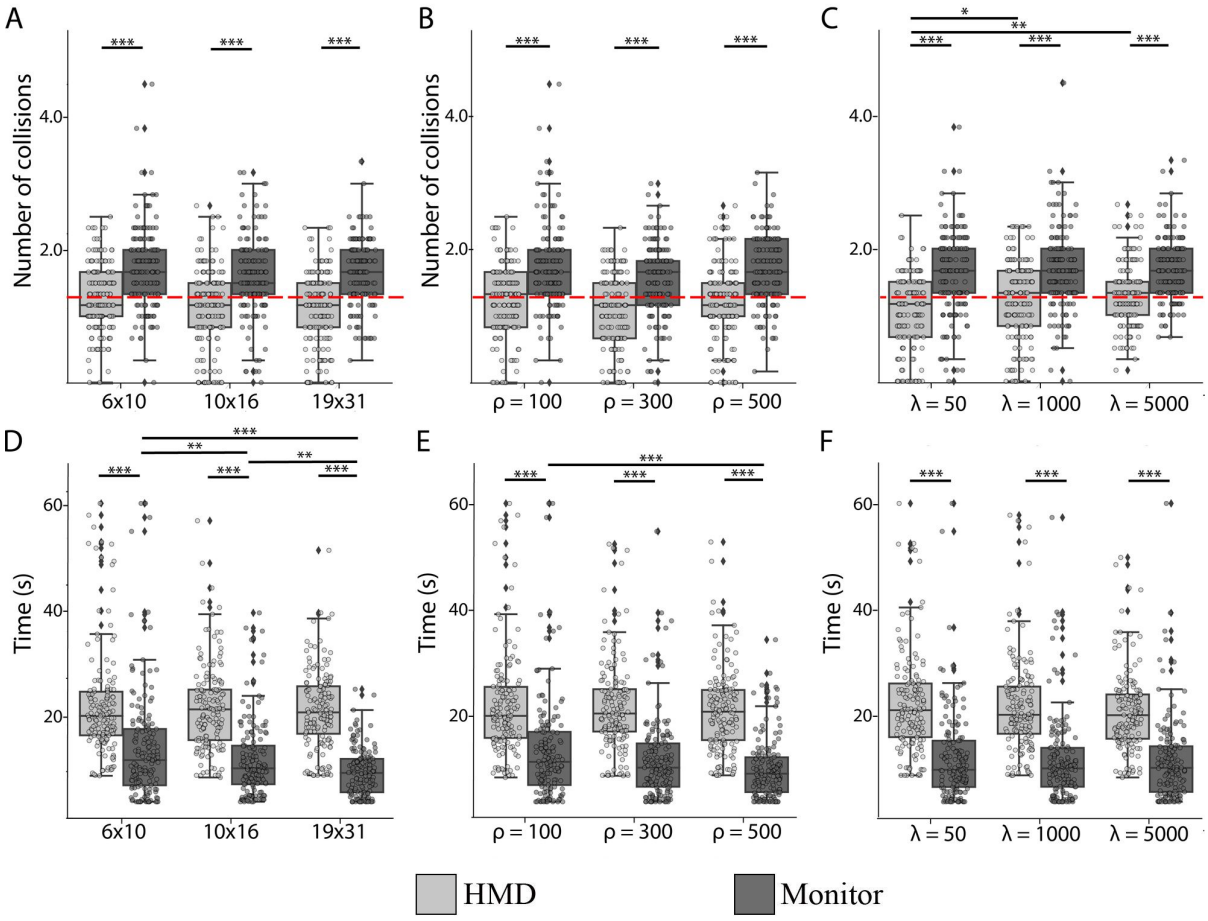


Figure 4.4: Obstacle avoidance. Data points represent each subject's average performance in a block with boxplots displaying median and interquartile ranges. ***Top:*** Average number of collisions across blocks for each subject within the condition specified by the x-axis. Red line represents chance level (1.25 collisions). ***Bottom***: Average time across blocks for each subject within the condition specified by the x-axis. Statistical significance was determined using ART ANOVA (*< .05, **< .01, ***< .001).

(Fig. 4.4, *Bottom*) averaged across repetitions in a block. Group F-values, along with their significance, are reported in Table 4.4.

Contrary to our expectations, neither the number of electrodes (Fig. 4.4A) nor phosphene size (Fig. 4.4B) had a significant effect on the number of collisions. Although the number of collisions decreased slightly with higher electrode counts (Table 4.3), this did not reach statistical significance. The only statistical differences

	Number of Collisions (± Std Dev)		Mean Time (s) (± Std Dev)	
	HMD	Monitor	HMD	Monitor
06x10 Array	1.279 (± 0.515)	1.734 (± 0.621)	22.138 (± 10.449)	14.045 (± 10.726)
10x16 Array	1.148 (± 0.602)	**1.603** (± 0.593)	21.483 (± 8.216)	11.622 (± 7.274)
19x31 Array	**1.117** (± 0.562)	1.663 (± 0.498)	**21.052** (± 7.388)	**9.234** (± 4.469)
ρ=100	1.253 (± 0.536)	1.739 (± 0.634)	22.000 (± 9.950)	13.378 (± 10.025)
ρ=300	**1.083** (± 0.558)	**1.553** (± 0.546)	21.699 (± 8.456)	11.594 (± 7.725)
ρ=500	1.208 (± 0.586)	1.709 (± 0.523)	**20.974** (± 7.797)	**9.929** (± 5.780)
λ=50	**1.037** (± 0.573)	**1.627** (± 0.637)	22.233 (± 9.523)	12.026 (± 8.412)
λ=1000	1.209 (± 0.610)	1.686 (± 0.607)	21.711 (± 9.004)	11.495 (± 8.183)
λ=5000	1.297 (± 0.473)	1.687 (± 0.466)	**20.728** (± 7.677)	**11.379** (± 7.850)

Table 4.3: Obstacle avoidance task: Average performance and time per trial across conditions. Best performances (lowest number of collisions/lowest time) for each grouping are presented in bold.

could be found between the scoreboard model (λ=50 µm) and axon map models (λ={100, 300}µm) for the HMD-based version of the task. However, participants performed around chance levels in all tested conditions.

The time analysis revealed a downward trend in time (better performance) with higher electrode counts, but only among the groupings in the monitor version. This trend in time reached significance for all comparisons within the monitor version (Fig. 4.4D). Similarly to comparisons across groupings of ρ values, there was a slight downward trend across the median time taken as phosphene distortion increased (Fig. 4.4, *F*).

A comparison between the two different versions of the task showed a clear difference in performance, with performance for the HMD version being drastically higher than the monitor version of the task. This trend reached significance across any grouping of device, ρ, or λ (Fig. 4.4, *Top*). There was also a difference in time taken between the versions of the task, with the HMD version taking longer for all groupings (Fig. 4.4, *Bottom*).

4.4 Discussion

4.4.1 Using an HMD May Benefit Behavioral Performance

The present study provides the first side-by-side comparison between HMD and monitor versions of different behavioral tasks using SPV. Importantly, we used a psychophysically validated SPV model to explore the expected behavioral performance of bionic eye users, for current as well as potential near-future devices, and found that participants performed significantly better in the HMD version than the monitor version for both tasks.

In the letter recognition task, participants achieved a higher mean F1 score across all conditions (Table 4.1). However, this trend was only significant for the hypothetical future devices and smaller phosphene sizes and elongations (Fig. 4.3, *Top*). While average time per trial was faster across all conditions with the HMD, the effect was not

	Num Collisions		**Time**	
	F	Signif.	F	Signif.
device	4.7538	8.85E-03	7.2265	2.51E-11
ρ	9.2904	1.02E-04	25.1790	2.00E-04
λ	4.8301	8.21E-03	19.8199	8.42E-33
display	207.3125	3.27E-42	335.6442	2.24E-01
device : ρ	1.2885	2.73E-01	3.6222	7.09E-01
device : λ	1.2039	3.08E-01	3.5733	1.67E-04
ρ : λ	0.2015	9.38E-01	1.1654	6.94E-01
device : display	1.3595	2.57E-01	6.5119	2.55E-01
ρ : display	0.3381	7.13E-01	0.9380	5.72E-01
λ : display	3.8149	2.24E-02	9.0722	6.27E-03
device : ρ : λ	1.0423	4.02E-01	3.1542	1.38E-02
device : ρ : display	0.9071	4.59E-01	1.9217	8.71E-01
device : λ : display	0.6814	6.05E-01	1.2380	9.63E-01
ρ : λ : display	1.5045	1.99E-01	2.0618	6.25E-01
device : ρ : λ : disp	0.9815	4.49E-01	2.2511	7.28E-01

Table 4.4: Obstacle avoidance task: F-value table for Aligned Rank Transform (ART) ANOVA. Values were calculated with the ARTool software package. "device" refers to the three simulated electrode grids, while "display"refers to the use of an HMD or monitor.

significant (Fig. 4.1, *Bottom*).

The difference in performance was even more evident in the obstacle avoidance task, where performance (as measured by number of collisions) for the HMD version was significantly higher than the monitor version across all conditions (Fig. 4.4, *Top*). It is also worth pointing out that participants were able to complete the task faster with higher electrode counts in the monitor-based version of the task. Since the walking speed was fixed across all conditions, this likely indicates that the task was easier with higher electrode counts.

Overall these results suggest that participants were able to benefit from vestibular and proprioceptive cues provided by head movements and locomotion during the HMD version of the task, which is something that is available to real bionic eye users but cannot be replicated by a mouse and keyboard.

4.4.2 Increased Phosphene Elongation May Impede Performance

Whereas previous studies treated phosphenes as small, discrete light sources, here we systematically evaluated perceptual performance across a wide range of common phosphene sizes (ρ) and elongations (λ). As expected, participants performed best when phosphenes were circular (scoreboard model: $\lambda = 50\,\mu m$; Tables 4.1 and 4.3), and increasing phosphene elongation (λ) negatively affected performance.

In the letter recognition task, participants using the scoreboard model (λ=50 µm) achieved a perfect median F1 score of 1.0 (Fig. 4.3C), which is much better than the behavioral metrics reported with real Argus II patients Cruz et al. (2013). Conversely, performance approached chance levels when increasing λ to 5000 µm.

In the obstacle avoidance task, the only significant findings within one version of the

experiment were between the scoreboard model ($\lambda = 50\,\mu m$) and either of the larger λ values. This suggests that elongated phosphenes make obstacle avoidance more challenging than the scoreboard model. However, participants performed around chance levels in all tested conditions, which was also true for real Argus II patients He et al. (2020).

Contrary to our expectations, phosphene size (ρ) did not systematically affect performance (Fig. 4.3B, Fig. 4.4B). The best performance was typically achieved with $\rho = 300\,\mu m$. This is in contrast to previous literature suggesting smaller phosphene size is directly correlated with higher visual acuity Chen et al. (2009); Han et al. (2021) Overall these findings suggest that behavioral performance may vary drastically depending on the choices of ρ and λ. This is important for predicting visual outcomes, because ρ and λ have been shown to vary drastically across bionic eye users Beyeler et al. (2019b), suggesting future work should seek to use psychophysically validated SPV models when making theoretical predictions about device performance.

4.4.3 Increasing the Number of Electrodes Does Not Necessarily Improve Performance

As expected, letter recognition performance improved as the size of the electrode grid (and therefore the FOV) was increased from 6×10 to 10×16 and 19×31 (Fig. 4.3A). This performance benefit was also observed in the time it took participants to recognize the letter (Fig. 4.3D), and is consistent with previous literature on object recognition Thorn et al. (2020).

However, electrode count did not affect behavioral performance in the obstacle avoidance task. Whereas there was a slight increase in performance scores for devices with more electrodes (Fig. 4.4A), this effect did not reach significance.

Overall these results are consistent with previous literature suggesting that, for most tasks, the number of electrodes may not be the limiting factor in retinal implants Beyeler et al. (2017); Behrend et al. (2011).

4.4.4 Limitations and Future Work

Although the present study addressed previously unanswered questions about SPV, there are a number of limitations that should be addressed in future work as outlined below.

First, in an effort to focus on the impact of phosphene size and elongation on perceptual performance, we limited ourselves to modeling spatial distortions. However, retinal implants are known for causing temporal distortions as well, such as flicker and fading, which may further limit the perceptual performance of participants Beyeler et al. (2017).

Second, the displayed stimuli were not contingent on the user's eye movements. Even though current retinal implants ignore eye movements as well, there is a not-so-subtle difference between a real retinal implant and a simulated one. Since the real device is implanted on the retinal surface, it will always stimulate the same neurons, and thus produce vision in the same location in the visual field—no matter the eye position. This can be very disorienting for a real patient as shifting your gaze to the left would not shift the vision generated by the implant. In contrast, a participant in a VR study is free to explore the presented visual stimuli with their gaze, thus artificially increasing the FOV from that offered by the simulated device. Consequently, the here presented performance predictions may still be too optimistic. In the future, simulations should make use of eye tracking technologies to update the scene in a gaze-contingent way.

Third, we did not explicitly measure the level of immersion across the two display types

(HMD and monitor). Instead, we assumed that viewing a scene that updates with the user's head movement through an HMD would lead to a higher level of immersion. Although this may be true for realistic virtual environments Miller and Bugnariu (2016), this has yet to be demonstrated for SPV studies. Future SPV work should therefore explicitly measure the level of immersion and/or a user's sense of presence.

Fourth, the obstacle avoidance task did not have a meaningful time metric. Although participants performed the task significantly faster in the monitor-based version, this is likely an artifact due to the walking speed of participants not being consistent between versions of the task. Participants moved much slower with the HMD as they were not able to see the real world around them. Future studies should take this into consideration and correct for each participant's walking speed within desktop versions of tasks.

Fifth, the study was performed on sighted graduate students readily available at the University of California, Santa Barbara. Their age, navigational affordances, and experience with low vision may therefore be drastically different from real bionic eye users, who tend to not only be older and prolific cane users but also receive extensive vision rehabilitation training.

Interestingly, we found vast individual differences across the two tasks (individual data points in Figs. 4.3 and 4.4) which were not unlike those reported in the literature Cruz et al. (2013); He et al. (2020). Subjects who did well in one experiment tended to do well across all versions of both experiments (data not shown), suggesting that some people were inherently better at adapting to prosthetic vision than others. Future work should therefore zero in on the possible causes of these individual differences and compare them to real bionic eye users. Studying these differences could identify training protocols to enhance the ability of all device users.

Chapter 5

Optimizing Prosthetic Vision: The Impact of Raster Patterns on Visual Task Performance and User Experience

While high-resolution visual prostheses are in development with hundreds or even thousands of stimulating electrodes, charge density limits prevent simultaneous activation of large numbers of electrodes. Additionally, continuous stimulation desensitizes the underlying tissue to subsequent stimulation (Fornos et al., 2012). The current solution to address these occurrences is electrode rasterization, the process of only activating subsets of electrodes in rapid succession (see Figure 3.2).

Many studies have focused on modeling phosphene shape, but only a few have attempted to model temporal distortions. Two notable examples are Avraham et al. (2021) and Thorn et al. (2020), who introduced a temporal model to capture the clinically reported phenomena of phosphene fading (where continued stimulation of the

same electrode would lead to the elicited phosphene to fade out over the timecourse of 500 ms to 2 s) and phosphene persistence (where patients would continue to see a phosphene even seconds after stimulus offset).

However, no simulation study has so far considered the perceptual consequences of electrode rasterization. The patterns used in electrode rasterization are often seemingly random, or simply straight lines (Second Sight (2013)). While the rasterization rate is often set higher than what is theoretically perceivable by human, patients often describe what they're seeing as fireworks or sparkles. It is conceivable that activating only a subgroup of electrodes would reduce the quality of prosthetic vision, by temporally segmenting the visual stimulus into separated, incoherent parts. It is also conceivable that directional rasterization of electrodes (e.g. columns going left to right or rows moving top to bottom) would introduce apparent motion, which could interfere with a patient's ability to correctly judge the direction of motion of an object.

To assess the impact of electrode rasterization on perceptual performance, we designed an SPV experiment for two multi-alternative forced choice (MAFC) tasks that are commonly employed to evaluate visual prostheses: letter and motion discrimination. By systematically evaluating raster patterns, we expected to find certain biases and strategies that were dependent on the electrode rasterization pattern. We aimed to identify patterns associated with high performance under different raster settings, along with raster settings that may perform better for certain tasks. The outcomes of our research may not only inform the design and optimization of retinal prostheses but also contribute to a more comprehensive understanding of the factors influencing the effectiveness of these systems.

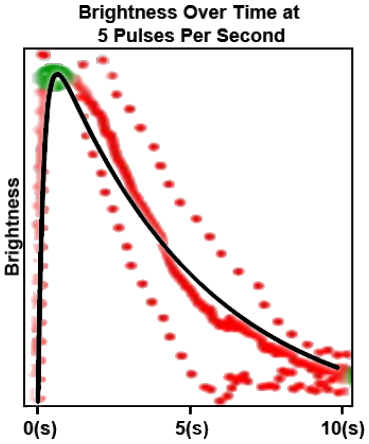
5.1 Methodology

5.1.1 Participants

We recruited 48 participants with normal or corrected-to-normal vision from the Department of Psychological & Brain Sciences research participant pool at the University of California, Santa Barbara (UCSB). Participants served as 'virtual patients', with ages ranging from 18 to 29 years ($M = 19.71$, $SD = 2.72$ years), and 20 participants identifying as male (28 female). Among these participants, 10 had never used VR before, 33 had used VR 1–5 times, 3 had used it 5–10 times, 1 had used it 10–20 times, and 1 had used it 20+ times. Potential participants were excluded if they reported being prone to cybersickness or sensitive to flashing lights. The study was approved by UCSB's Institutional Review Board.

5.1.2 Simulated Prosthetic Vision

As in the previous chapters, we used BionicVisionXR to generate realistic estimates of prosthetic vision. BionicVisionXR was processed in real time by a desktop PC (Intel i9-9900k, 16GB DDR4 memory with an Nvidia RTX2070Su) and wirelessly sent to a head-mounted display (HMD) (HTC VIVE Pro Eye with wireless adapter, HTC Corporation). For this study, we simulated a retinal implant composed of a 10×10 electrode array centered over the fovea. The electrodes were simulated as points with an electrode-to-electrode spacing of 400 µm. The spatial distortion parameters (ρ and λ varied throughout the experiment (see Section 5.1.5), but all trials used gaze congruent rendering (see Section 3.4).



Our temporal model had three open parameters (see Eqs. 3.3–3.4): brightness decay (τ_b), desensitization decay (τ_n), and a scaling factor (α). Values for these parameters were chosen by fitting Fig. 4, subject 5 of Fornos et al. (2012). The red curve represents the subject average, with red dots showing the range across trials. The black curve represents the temporal fading model implemented across pixels.

Figure 5.1: **Temporal Brightness Curve**

To model temporal dynamics, we used BionicVisionXR's built in temporal processing (see Section 3.2.2). We used patient data from Fornos et al. (2012) to determine the three open parameters from equations 3.3 and 3.4: $\tau_b = 5s$, $\tau_n = 0.2s$, $\alpha = 0.2$

5.1.3 Raster Strategies

In actual prosthetic implementations, a process known as "rasterization" is employed, whereby electrodes are stimulated in disparate timing groups. This method ensures that only a subset of electrodes is stimulated concurrently, and the device swiftly cycles through different groups. The advantages of this approach include mitigating tissue damage risks, prolonging battery life, and minimizing neural desensitization. Horsager et al. (2010) highlighted the importance of understanding these spatiotemporal effects. Even after controlling for electric field interactions, they found that subjects could discern phase differences across electrodes as small as 3ms.

In order to maintain user comfort, a delicate equilibrium typically exists between the quantity of timing groups and the frequency of stimulation (Second Sight (2013). In this study, we confined our efforts to five raster groups, each receiving stimulation at 4.545 Hz (220ms per pulse train, see Fig 5.2). This frequency closely approximates the

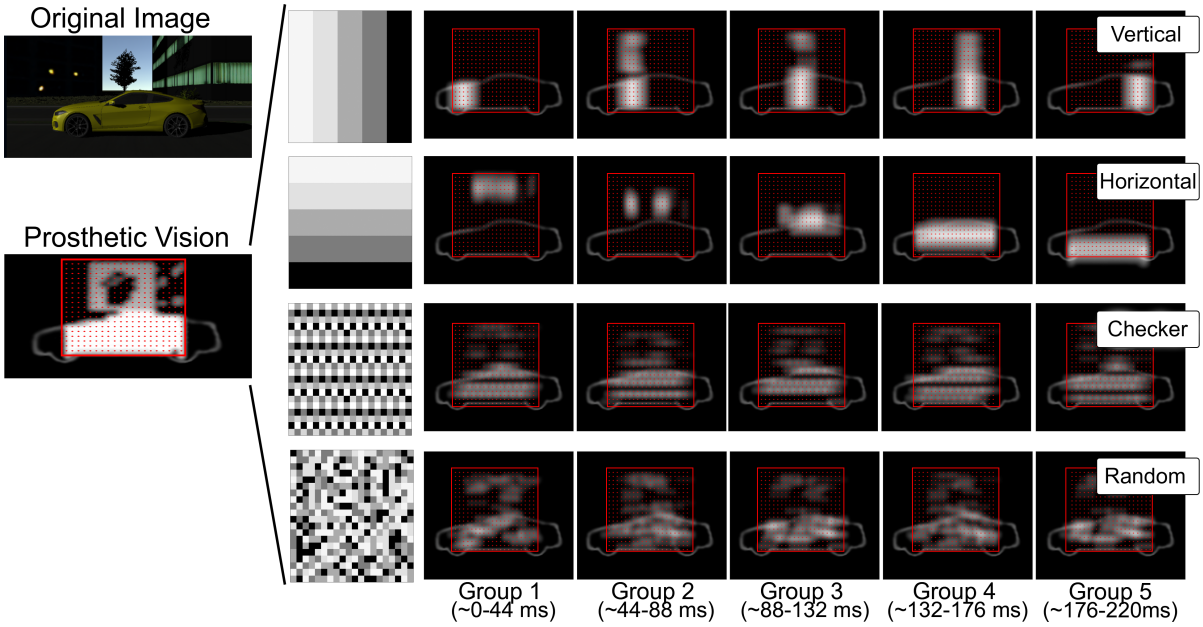


Figure 5.2: **Raster Patterns: Left -** The image is converted into an electrode activation pattern and used to simulate prosthetic vision. Red dots represent the locations of simulated electrodes. **Right -** To avoid high levels of current, only 20% of the electrodes are active concurrently. The four different raster patterns shown were assessed in the behavioral tasks. Vertical lines moving left to right, horizontal lines moving top to bottom, equally spaced electrodes in a checker pattern, and random groups are re-randomized every five frames. The groups were rendered as close to 5Hz as possible with the headset's 11ms frame timing.

5Hz timing used in the temporal model, given the headset's rendering is fixed at 90Hz/11ms.

The selection of five groups facilitates equal group sizes while maintaining an inter-pulse delay between timing groups of 44ms (as detailed in Fig 5.2). This delay duration aligns with previously published inter-pulse delays of 25-83ms (Yücel et al., 2022) and 50ms (Christie et al., 2022).

We conducted an analysis of four unique raster strategies, all readily adaptable to current device architectures (see Fig. 5.2):

- **Horizontal:** Each group contains 2 adjacent rows of 10, starting with the top two rows and moving down to the next 2 rows on each subsequent frame.
- **Vertical:** Each group contains 2 adjacent columns of 10, starting with the left two columns and moving right to the next 2 columns on each subsequent frame.
- **Checkerboard:** Maximizes the spacing between electrodes for each frame creating a "checkerboard" pattern. Subsequent frames maintain the checker pattern but are shifted to create the maximum distance between active electrodes on two adjacent frames (See B.1)
- **Random:** Creates randomized groups for every sequence of five frames, insuring all electrodes are active one time during the sequence.

Timelapse video files showcasing the various conditions and raster patterns are also available on the paper's GitHub page:

5.1.4 Tasks

To investigate the effects of rasterization patterns, we simulated two tasks, both involving eight-alternative forced choice tests. In both tasks, a 5 s video stimulus played on a simulated computer monitor within a darkened virtual room. The goal of both tasks was to identify the stimulus being shown. After the video was played, participants used a joystick to point in one of eight available directions (up, down, left, right, up/left, up/right, down/left, and down/right). Moving the joystick would show the selected response and the trigger button was used to finalize the selection. While participants had unlimited time to respond, they were instructed not to spend more

than 5–10 seconds making each decision. Throughout the entire experiment, participants received feedback for every trial, indicating whether they were correct and, if not, what the correct response was.

- **Letter identification:** Participants were presented with one of eight Snellen optotypes (C, D, E, F, L, O, P, and T). This study replicated a previous experiment conducted with participants using visual prostheses (Cruz et al., 2013) and later with simulated prosthetic vision (Kasowski and Beyeler, 2022). In this task, a letter was displayed on the virtual monitor for five seconds. Participants were asked to identify the letter shown, and each joystick direction corresponded with selecting a specific letter. For example, pushing the joystick forward and left would show the letter "C", which could then be selected by pushing the joystick trigger button.

- **Motion discrimination:** Participants were presented with stimulus videos created using pulse2percept (Beyeler et al., 2019b). These videos were five seconds long and featured a grating of bars moving perpendicular to the bars' orientation. Participants were asked to identify the direction of motion, and each joystick direction corresponded with each direction. For example, pushing the joystick forward and left would show an arrow pointing up/left, which could then be selected by pushing the joystick trigger button.

5.1.5 Procedure

Before arriving, participants completed surveys to determine eligibility and collect demographic information. Sessions began with participants reading instructions that included tips for using bionic vision (see supplemental material on GitHub). Subsequently, they put on the headset and completed HTC's Eye Calibration procedure

(included with 'VIVE Console'). To ensure the eye-tracker's accuracy, a subset of participants ($n = 30$) also completed an additional "follow the dot" task (see Appendix A).

Participants started with one of the two tasks (counterbalanced). To promote learning, we employed increasing spatial distortion values as participants progressed through training. Training started with two normal-vision (non-SPV) trials, followed by three sets of five trials with incrementally increasing difficulty levels ($\lambda = 50, 50, 1000$; $\rho = 150, 300, 300$). The first set used low distortion values ($\lambda = 50$, $\rho = 150$) and 400 (20x20) electrodes spaced 300 µm apart. This gave participants a pixel-like representation of the image with a large field of view. The second set used slightly higher radial distortion ($\rho = 300$) and 100 (10×10) electrodes spaced 400 µm. This introduced some blurring of the pixels and restricted the field of view to $\sim 14.6^o$x14.6^o. This forced participants to scan the screen with head/eye movements during the letter task and created a limited viewing window for the motion discrimination task. The third difficulty level falls within the realistic range of distortion values and electrode configurations for current epiretinal devices (Beyeler et al., 2019b). Extensive piloting demonstrated a 70-90% post-training accuracy range for the baseline condition (temporal effects but no electrode rasterization). The final difficulty level reached during training was applied to all trials in the experimental phase.

During learning, there were no temporal effects (including electrode rasterization), but the display was gaze-contingent (see Sec 3.4). After the initial training phase, every participant began with the 'No Raster' baseline condition. This introduced temporal effects (see Sec 3.2.2), but all electrodes were active on every frame. Participants then completed the four raster conditions in different, counterbalanced orders. Each condition (including 'No Raster') had 48 trials divided into 6 blocks of all 8 stimuli. The stimulus order was randomized after each set of 8, but participants were not told

they were completing blocks and likely assumed randomization. Upon completion of each condition, participants were asked to rate the difficulty of the condition on a 5-point Likert scale.

5.1.6 Data Collection and Analysis

For each trial, the participant's choice and the correct choice were recorded, along with their eye and head positions during the trial. A linear mixed-effect model (`mixedlm` from the Python package `statsmodel`) was used to account for using a within-subjects repeated measures design (as suggested by Magezi (2015)). To use a 'design-driven, maximal model' (Barr et al., 2013), our variables of interest ('Block', 'RasterStrategy', and 'Task') were modeled as fixed effects with full interactions. Data was grouped by subject, with 'Task' treated as a nested random effect. The variance/covariance structure was set with 'Block' and 'RasterStrategy', allowing the random variance of each to have varying slopes/intercepts across the different 'Task/Subject' combinations. The final equation was:

$$\begin{aligned} Accuracy &\sim Block * C(RasterStrategy) * C(Task) \\ vc\_formula &= \{1 + Block, 1 + C(RasterStrategy)\} \\ re\_formula &= 1 + C(Task) \\ groups &= C(Subject) \end{aligned} \tag{5.1}$$

To study pairwise comparisons for task accuracy and difficulty ratings, subsets of the dataset were created for each 'Task'/'Block' combination. Reported p-values were corrected for multiple comparisons using the Benjamini-Hochberg procedure. For the

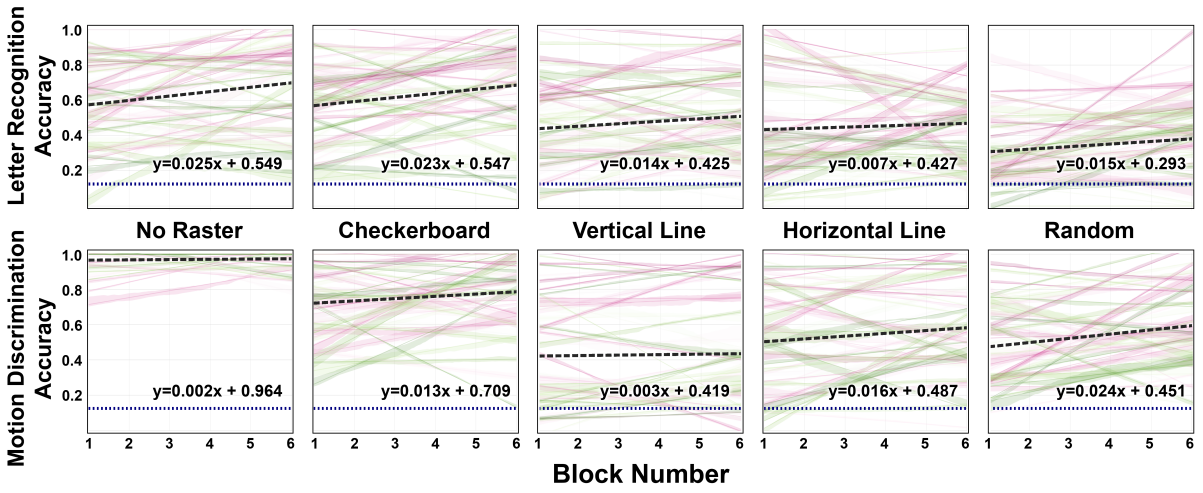


Figure 5.3: **Linear Mixed Effects Model-** Linear mixed effects models for letter recognition (top) and motion discrimination (bottom) over 6 blocks. The mixed effects model was a maximal model grouped by subject with a random effects covariance structure of ('Block', 'Raster Setting') across tasks. The full interactions of 'Block', 'Raster Setting', and 'Task' were modeled as fixed effects. Colored lines represent the fit for each individual subject, with the blue dashed line representing chance level.

accuracy analysis, a one-tailed t-test was performed to verify higher performance in the checkerboard condition. Conversely, in the difficulty rating analysis, a one-tailed t-test was used to verify lower difficulty ratings in the checkerboard condition.

To investigate whether better performers employed specific strategies, the head position and eye tracking information were used to visualize the locations participants focused on the most. Moreover, head and eye movement velocities were examined for correlations with performance. This analysis aimed to provide insights into the strategies that led to better task performance under different rasterization conditions, potentially informing future prosthetic vision designs and training methods for users.

5.2 Results

5.2.1 Linear Mixed Effect Model Results

The linear mixed effects model (see Fig. 5.3) had a much better fit (AIC= −832.82, BIC= −683.68) than null models independently examining the effects of block (AIC= 1425.53, BIC= 1443.42), raster setting (AIC= 566.64, BIC= 602.43), or task (AIC= 943.61, BIC= 961.51). The main effects of 'Task', 'RasterStrategy', and 'Block' were significant, allowing for subsequent post hoc pairwise comparisons. In the baseline model (block= 0, checkerboard rasterization, and the letter recognition task), participants performed at an average accuracy of 0.546 (95% CI = $[0.479, 0.614]$). Participants improved their accuracy scores at an average rate of 0.023 per block ($p < .001$, 95% CI = $[0.010, 0.036]$) and had higher accuracy scores on the motion task ($p < .001$, 95% CI = $[0.083, 0.242]$). There was a likely negligible difference in the letter recognition task for the 'no raster' condition ($p = 0.935$, 95% CI = $[-0.072, 0.078]$) and participants had accuracy scores that were significantly lower when the rasterization pattern was vertical lines ($p < 0.01$, 95% CI = $[-0.197, -0.046]$), horizontal lines ($p < 0.01$, 95% CI = $[-0.195, -0.045]$), or random ($p < .001$, 95% CI = $[-0.328, -0.177]$).

Furthermore, the interaction between raster strategy and task was significant for all raster settings except random, indicating that the effect of these raster settings is influenced by the task. Specifically, the differences with the checkerboard setting were more pronounced, with the 'No Raster' condition yielding higher scores ($p < .001$, 95% CI= $[0.155, 0.350]$), and lower scores observed for vertical lines ($p = .001$, 95% CI= $[-0.265, -0.071]$) and horizontal lines ($p = .040$, 95% CI= $[-0.199, 0.004]$).

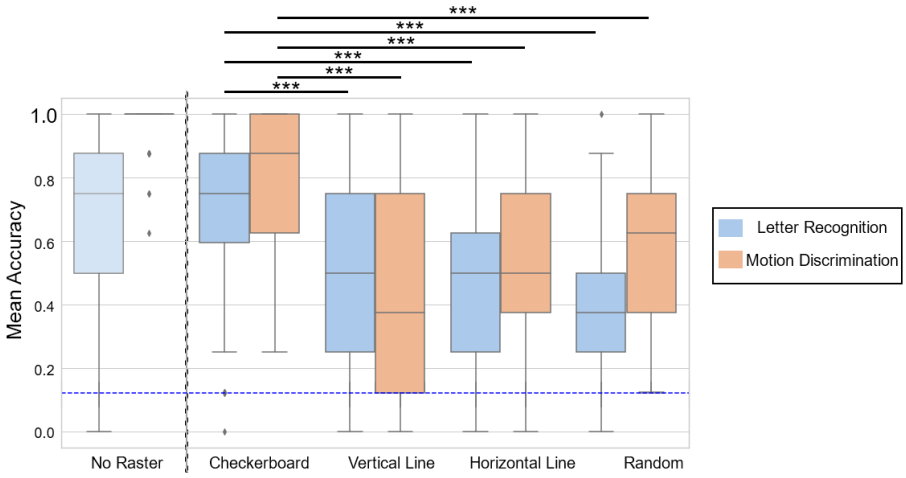


Figure 5.4: **Rasterization Results -** Final block (Block 6) accuracy results for both the letter recognition and motion discrimination tasks. Significant performance improvements are evident in the "Checkerboard" condition compared to the other three raster patterns. Blue dashed line represents chance level (1/8). Significance determined with one-tailed pairwise t-tests (Benjamini/Hochberg FDR correction for multiple comparisons, $***= p < .001$)

5.2.2 Checkerboard Rasterization Outperforms Similar Complexity Patterns

In both tasks, performance with the 'Checkerboard' raster setting was significantly higher than all other raster settings across all blocks. By block number six, participants were able to consistently score above chance levels with the checkerboard raster pattern while some participants still performed below chance levels for the other conditions. Figure 5.4 shows the accuracy across both tasks (letter recognition=LR, motion discrimination=MD) for no rasterization (LR=0.703±0.269, MD=0.969±0.079), checkerboard (LR=0.695±0.260, MD=0.786±0.213), vertical lines (LR=0.484±0.278, MD=0.450±0.305), horizontal lines (LR=0.471±0.248, MD=0.563±0.265), and random (LR=0.393±0.246, MD=0.591±0.305).

It is important to note the possibility of an artifact where the 'Checkerboard' performance matches the 'No Raster' condition in the letter recognition task. While

'Checkerboard' performance is significantly higher than the other conditions, it is likely still not as high as the performance in the 'No Raster' condition. However, every subject started with the 'No Raster' baseline condition and would have likely performed better had they encountered the condition later in the experiment. This organization was intentional, as many subjects could not perform the other conditions without first being trained without rasterization.

5.2.3 Performance Improvement Evident Across Raster Conditions

The maximal mixed effect model revealed no random variance could be attributed to the 'Block' interactions, while the fixed effect of 'Block' was significant. This implies 'Block' significantly affects performance across all 'Task' and 'RasterStrategy', but the variance in the effect is not significant across different 'RasterStrategy' and 'Task' pairings. As a result, removing 'Block' interactions created a slightly better fitting model (AIC= -907.25, BIC= -811.80) with a fixed slope regardless of 'RasterStrategy' or 'Task' ($m = 0.014$).

Although not statistically significant, the effect of 'Block' in the maximal model (and its interactions) may better represent the reality of subject learning throughout the different conditions (see Fig. 5.3). For instance, the model without 'Block' interactions would suggest subjects could perform above 100% accuracy by Block #6 in the 'No Raster' motion task.

5.2.4 Self-Reported Difficulty Ratings

Figure 5.5 displays the participant self-reported difficulty ratings for both tasks. The 'Checkerboard' condition was rated significantly lower in difficulty compared to the

other three conditions across both tasks. The difficulty ratings have an inverse relationship with performance measurements for each condition (i.e. participants scored higher in tasks rated lower in difficulty). In the letter recognition task, the 'No Raster' and 'Checkerboard' conditions had similar, lower difficulty scores compared to the other three conditions. For the motion task, the 'No Raster' condition was rated considerably lower in difficulty than the 'Checkerboard' condition. However, the 'Checkerboard' condition was still rated as significantly less difficult than the other three raster conditions.

5.2.5 Systematic Stimulus Dependent Biases in Participant Responses

For both tasks, participants' selections were anticipated to be biased for certain stimuli or raster conditions. In the letter recognition task, we expected participants to select letters that closely resembled the presented stimulus, particularly confusing letters

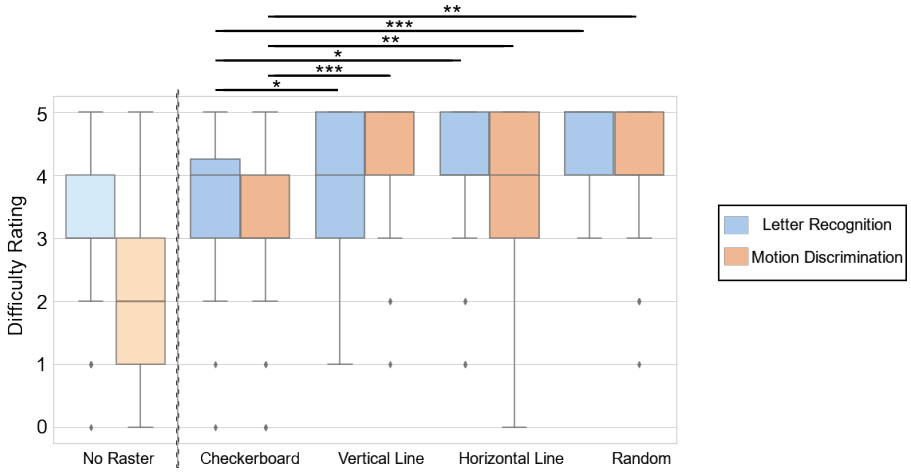


Figure 5.5: **Difficulty Ratings.** Distribution of difficulty ratings completed at the end of each condition for Letter Recognition (Left-side bars) and Motion Discrimination (Right-side bars). Significance bars determined with one-tailed pairwise t-tests (Benjamini/Hochberg FDR correction for multiple comparisons, $* = p < .05$, $** = p < .01$, $*** = p < .001$)

within groups [C, D, O] and [E, F, P]. In the motion discrimination task, it was expected that the raster setting would influence participants' choices, with selections likely biased toward the apparent motion in the vertical line (left to right) and horizontal line (top to bottom) conditions. Participant responses to each stimulus can be seen in Figure 5.6.

In the letter recognition tasks, there was some bias towards similar-looking letters, but the strongest biases existed towards the letters 'D' and 'P'. This is possibly due to those being the options when the joystick was pressed forward or pulled backwards. Subjects may have felt those joystick movements were the most natural and subconsciously favored them when randomly guessing. There was no apparent

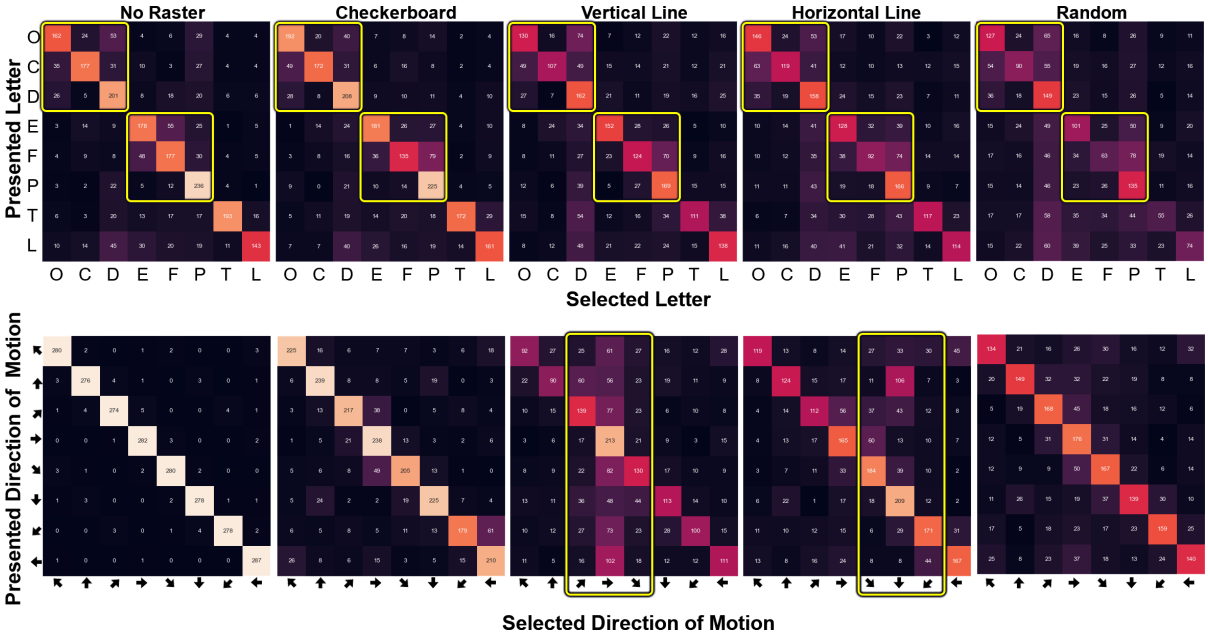


Figure 5.6: **Stimulus Response Matrix -** The figure illustrates the relationship between the selected response (x-axis) and the specific stimulus presented (y-axis). In the letter recognition task (Top), there were no appreciable effects observed when altering raster modes. Participants did show slightly elevated biases towards letters with similar properties ([C,D, and O] and [E, F, and P]). Meanwhile, in the motion discrimination task (Bottom), responses tended to favor the direction of apparent motion for both vertical and horizontal lines. Yellow rectangles represent the expected biases.

differences in bias between the different raster patterns, with the majority of incorrect responses being in the expected groupings or 'D'/'P' across conditions.
The results for the motion discrimination task matched our hypothesis, with performance being lowest for the vertical and horizontal line conditions. Biases in these conditions matched the apparent direction of motion for the moving bars and are highlighted with a yellow rectangle in Figure 5.6. Notably, these systematic biases do not exist in the checkerboard or random conditions.

5.3 Discussion

5.3.1 Implications of the Results

This investigation suggests the 'Checkerboard' rasterization pattern outperforms other patterns in both letter recognition and motion discrimination tasks. These results echo prior studies indicating complexities arise from merging nearby phosphenes Rizzo et al. (2014); Horsager et al. (2010); Wilke et al. (2011); Bosking et al. (2017), especially when deciphering intricate shapes like letters (Christie et al., 2022). Sequential patterns have improved performance with cortical prostheses (Beauchamp et al., 2020; Oswalt et al., 2021), but sequential stimulation alone has failed to improve performance for recognizing complex shapes with retinal stimulation (Christie et al., 2022). The current work suggests careful selection of timing groups is critical in facilitating functional vision with these devices. The current work has also shown that rasterization is a plausible way to introduce motion effects. While this could be used advantageously if the direction of apparent motion matches motion in real life (e.g. lines moving to the right as an object moves to t he right), the checkerboard rasterization performed higher even when this was the case. Additionally, decreased task difficulty ratings were

correlated with higher performance, reiterating the need for a user-centered approach in designing visual accessibility devices (as previously reviewed in Kasowski et al., 2023). User insights can assist in identifying raster patterns that not only enhance performance but also streamline tasks and reduce perceived difficulty. Although performance metrics hold crucial importance, alleviating task difficulty can prevent mental fatigue and foster higher rates of device adoption.

5.3.2 Gaze Strategies and Performance

The analysis of gaze strategies revealed high performers adopted specific strategies depending on the raster setting. In the letter recognition task, all three performance groups focused on the center of the letter and made small movements in all directions when presented with the 'No Raster' and 'Checkerboard' conditions. Interestingly, high performers in our study applied different strategies for the other raster settings, focusing on the bottom and top-right parts of the letter. These specific letter areas contain the most information and can be used to discriminate between similar letters. These gaze patterns were not observed in the mid or low performing groups and may partially explain the discrepancy The fact that top performers were able to adjust their gaze strategies depending on the raster setting highlights the importance of training and user adaptation in prosthetic vision devices.

5.3.3 Insights for Prosthetic Vision Development

Our results provide valuable insights for the development of prosthetic vision systems. The improved performance observed with the checkerboard raster setting suggests incorporating this pattern into prosthetic vision devices may enhance the user experience and facilitate the processing of visual information. Furthermore, the

differences in gaze strategies among high, middle, and low performers indicate that training and user adaptation play a crucial role in optimizing the performance of prosthetic vision systems. Future research should investigate the long-term effects of gaze training on the performance of prosthetic vision devices. Studies exploring the neural mechanisms underlying the processing of different raster patterns, as well as the development of adaptive rasterization algorithms that adjust to the user's gaze strategy, may also contribute to the advancement of prosthetic vision technology.

5.3.4 Limitations and Future Directions

While our study provides valuable insights into the effects of raster patterns on task performance and difficulty ratings, some limitations should be considered. First, the sample size of the study was relatively small, which may limit the generalizability of the findings. Additionally, we focused on healthy participants with normal vision and assumed all electrodes elicited percepts (which is not always the case in real devices). While this approach allowed for controlled comparisons of rasterization patterns, it is possible that individuals with visual impairments would perform differently or adopt different strategies when using prosthetic vision devices. Future research should involve a larger and more diverse participant pool to confirm the robustness of the results and ideally compare performance to real prosthesis users.

Second, our study focused on two specific tasks: letter recognition and motion discrimination. Additional research examining the effects of raster patterns on a wider range of tasks, including object recognition, facial recognition, and navigation, would provide a more comprehensive understanding of the role raster patterns play in prosthetic vision systems.

Third, our study did not explore the impact of different electrode densities on task

performance and difficulty ratings. Investigating the interaction between raster patterns and electrode densities in prosthetic vision devices may further inform the design and optimization of these systems. Similarly, our simulation's temporal effects are based on the perceptual experience of prosthesis users. Our model simplifies electrode charge injection as a Gaussian decay away from the electrodes. In reality, the electrical properties of neural prostheses are complicated and not fully understood. The interaction between electrodes is not always predictable, especially in epiretinal devices (Christie et al., 2022). As with all simulation findings, results should be compared to the behavioral performance of real prosthesis users.

Lastly, our study (and previous simulation studies, see Dagnelie et al. (2007); Kasowski and Beyeler (2022)) revealed a large degree of individual difference in task performance. The current work examined gaze patterns of high performers, but there are likely other factors contributing to the large degree of difference in performance. In real prostheses, individual differences are often attributed to the placement of the device or differences in neuroanatomy. However, in simulation tasks all subjects are experiencing the same degree of distortion, but some participants consistently perform much higher than others. Identifying factors influencing performance outside of physical constraints may provide insights into training paradigms that result in more usability from current devices.

In conclusion, our findings demonstrate the potential benefits of incorporating checkerboard raster patterns into the design of retinal prosthetic devices. The superior performance and lower difficulty ratings associated with the checkerboard raster setting across both tasks suggest this pattern may be better suited for optimizing prosthetic vision in users. Furthermore, the identification of specific gaze and head movement strategies that enhance performance under different rasterization conditions could inform the development of targeted training programs for individuals using retinal

prostheses. Ultimately, these insights could lead to improved functional outcomes for people with visual impairments, enabling them to better navigate and engage with their environments through the use of advanced prosthetic vision technology.

Chapter 6

Enhancing Visual Prosthetics with AI: Investigating Smart Rasterization and Edge Detection for Improved Wayfinding Performance

Visual prosthetics offer an opportunity to tap into the existing neural circuitry of people who are blind and augment their visual senses like Google Glass or Microsoft HoloLens. With the low resolution and high distortion levels of current devices, it is necessary to simplify the visual scene before it is displayed. Most visual prostheses are equipped with an external VPU that can apply simple image processing techniques to the video feed in real time. Edge detection and contrast maximization are already routinely used in current devices. In the near future, these techniques may include computer vision-based algorithms aimed at improving a patient's scene understanding

Beyeler and Sanchez-Garcia (2022).

SPV studies suggest that one benefit of computer vision may be to provide an importance mapping that can aid scene understanding; that is, to enhance certain image features or regions of interest, at the expense of discarding less important or distracting information. This limited compensation may be significant to visual prosthesis patients carrying out visual tasks in daily life.

Based on this premise, researchers have developed various image optimization strategies, and assessed their performance by having sighted observers (i.e., *virtual patients*) conduct daily visual tasks under SPV (Boyle et al., 2008; Dagnelie et al., 2007; Al-Atabany et al., 2010; Li et al., 2018; McCarthy et al., 2015; Vergnieux et al., 2017). One of the most commonly explored strategies is to highlight visually salient information in the scene, which was able to improve eye-hand coordination (Li et al., 2017), obstacle avoidance (Stacey et al., 2011), object detection (Weiland et al., 2012; Han et al., 2021), and object recognition (Li et al., 2018; Wang et al., 2016). Another strategy is to use semantic segmentation highlight important objects Horne et al. (2016); Sanchez-Garcia et al. (2019); Han et al. (2021). These simulations allow a wide range of computer vision systems to be developed and tested without requiring implanted devices.

Along with most studies using inaccurate models of bionic vision, none of the simulations included temporal effects or raster patterns, which may further affect behavioral performance, as demonstrated in the previous chapter. It is therefore unclear how their findings would translate to real retinal prosthesis patients, whose phosphenes are large, elongated, and often fail to assemble into more complex percepts Rizzo et al. (2003); Wilke et al. (2011); Beyeler et al. (2019b); Erickson-Davis and Korzybska (2021).

To address these challenges, we sought to find way to use rasterization to not only limit

charge density and prevent desensitization, but also to aid the user in scene understanding. We hypothesized that by using temporal segmentation to present different objects at distinct times ('smart rasterization'), participants would more effectively recognize objects within each segment, leading to improved performance in a wayfinding task. Another 'smart' technique, 'smart edge detection', was used to show important structural edges, people, and moving targets while filtering out far away buildings or edges that would not contribute to scene understanding. This chapter will describe the methodologies and the two different 'smart' techniques that were found to improve performance in a simulated wayfinding task.

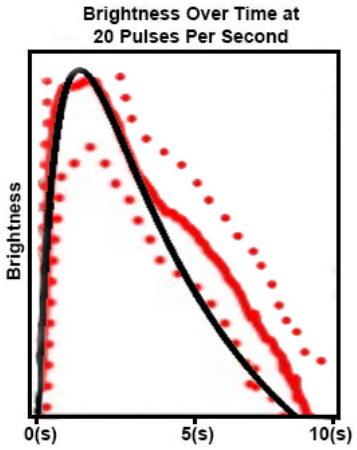
6.1 Methodology

6.1.1 Participants

Twenty four subjects with normal or corrected-to-normal vision were recruited and screened for sensitivity to flashing lights or motion sickness. Participants ranged from 18 to 40 years (25.04±5.72), with 10 participants identifying as male and 14 identifying as female. Among these participants, 5 had never previously used VR, 12 has used VR 1-5 times, 3 used VR 5-10 times, 2 used VR 10-20 times, and 2 previously used VR over 20 times. The study was approved by UCSB's Institutional Review Board.

6.1.2 Simulated Prosthetic Vision

BionicVisionXR was processed in real time by a desktop PC (Intel i9-11900k, 64GB DDR4 memory with an Nvidia RTX3090) and wirelessly sent to a head-mounted display (HMD) (HTC VIVE Pro Eye with wireless adapter, HTC Corporation). Spatial distortions were kept constant through the experiment, with $\rho = 200$ and $\lambda = 400$ (see



Our temporal model had three open parameters (see Eqs. 3.3–3.4): brightness decay (τ_b), desensitization decay (τ_n), and a scaling factor (α). Values for these parameters were chosen by fitting Fig. 3, subject 5 of Fornos et al. (2012). The red curve represents the subject average, with red dots showing the range across trials. The black curve represents the temporal fading model implemented across pixels.

Figure 6.1: **Temporal Brightness Curve**

Section 3.2.1). The temporal distortions were changed slightly from the last experiment and were matched to the same patient, but for a 20Hz stimulation (Fornos et al., 2012). We chose to match our temporal model to the 20Hz data since we had 5 raster groups and a 90Hz framerate. This effectively results in each electrode being stimulated at 18Hz. To match our temporal model to the data, we filled the open temporal parameters with $\tau_n d = 0.2$, $\tau_b = 1.5$, and $\alpha = .25$, resulting in phosphenes that would persist if the stimulus was removed prior to desensitization (see Section 3.2.2, Fig 6.1). As in the previous experiment, a gaze-congruent display was rendered by performing a double shift when processing the data (see Section 3.4)

Preprocessing Methods

Three different preprocessing protocols were investigated (see Figure 6.2):

1. 'Naïve Edge Detection' used a standard Sobel kernel, a common edge detection filter used in current devices that highlights region where there are rapid intensity changes, effectively emphasizing edges and transitions within an image.

2. 'Smart Edge Detection' effectively highlighted the outlines of people, bikes, and

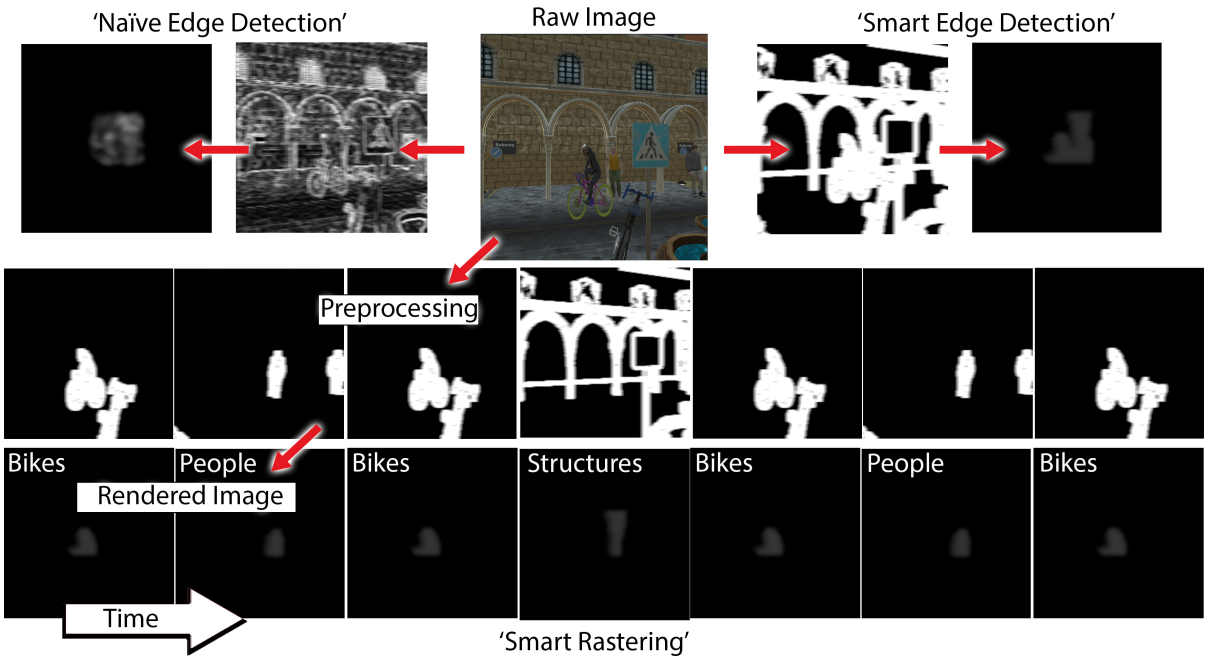


Figure 6.2: The experiment included three different pre-processing methods: 'Naive Edge Detection' (top left), 'Smart Edge Detection' (top right), and 'Smart Rastering' (bottom). The raw image underwent one of these three preprocessing methods before being rendered as prosthetic vision. For 'Naive Edge Detection', a standard 3×3 Sobel kernel was used. For 'Smart Edge Detection', the textures were removed from the objects and a 7×7 'enhancing kernel' was performed. This effectively removed background noise, leaving just people, bikes, and structural edges. For 'Smart Rastering', the image was segmented and presented in piece-by-piece in time, temporarily showing only bikes, only people, and only structural edges. The rastering occured in a way that bikes were shown most often, people second, and structural edges were shown the least. This was accomplished by presenting the groups as 'Bikes, People, Bikes, Structures, Bikes, People, Bikes' and repeating the pattern. This effectively showed people twice as often as structures, and bikes twice as often as people.

structural edges (e.g., of nearby buildings). While this may have seemed implausible years ago, technology and computer vision has advanced to a point where most modern smartphones are powerful enough to perform such algorithms.

3. 'Smart Rastering' separated the three semantic categories identified by 'Smart Edge Detection' in time, by displaying bikes, people, and structures in separate raster groups. Each group would display 'bikes only', 'people only', or 'structural edges' only. These groups would be displayed for all the frames within a 200 ms block (while continuing to also perform checkerboard rasterization). This timing was chosen after piloting and attempted to balance seeing the different groups often enough to be meaningful, while also showing enough of one block to be able to gather information about obstacle locations and speeds. Due to bikes being the more critical group, they were shown twice as often as people, and four times as often as structural edges. This was accomplished by displaying 'bikes, people, bikes, structural edges, bikes, people, bikes' and then repeating the pattern.

All three approaches used a checkerboard raster pattern, set at 90 Hz (to coincide with the framerate of the HMD). The 'Smart Rasterization' pattern was chosen based on our hypothesis, that adding information through temporal segmentation would aid in scene understanding and navigational ability. Part of this method included highlighting bikes, people, and structures in order to segment them. To insure any performance gains were from the temporal segmentation, the "Smart Edge Detection" acted as a control. It was simply "Smart Rasterization" without the temporal segmenting. However, this is not what is currently standard in visual prostheses which typically use simple edge detection algorithms and downscaling ('Naïve Edge Detection').

Figure 6.3: The entrance to the subway station. The participant always started in the same location, in front of the fountain (highlighted square). They were instructed to go to the left or right side of the station while avoiding collisions with people, objects, and bikes. A trial was deemed successful when the participant was inside the subway station and on the correct side. A trial ended upon success, when the participant collided with a biker, or when they ran out of time (45 seconds).

6.1.3 Task

The task performed was a wayfinding task with obstacle avoidance. This task was chosen because it requires participants to perform multiple simpler tasks in order to be successful. Participants needed to be able to identify the structures around them, perform motion discrimination, plan a path, and avoid the objects/people they were tasked with recognizing. Participants started on one side of a small town square and were instructed to go to the left or right side of the subway station. The subjects were given 45 seconds to complete the trial, and at 10 seconds a timer appeared on the screen counting down with the time remaining. The trial ended when the participant made it to the correct side of the subway station and past the arches (see Figure 6.3), ran out of time, or got hit by a bicycle. The participants were instructed to prioritize avoiding bicycles, and the trial would end while a crashing sound played. They were told that their second objective was making it to the subway station, and their third was to avoid as many other collisions as possible. The experiment was a within-subjects block design, with each preprocessing method being a complete 'block'. The block order was cross-balanced, and all blocks had the same set of fifteen trials (in a randomized order). Participants began by completing 5 rounds in a dark virtual space containing people, objects, and bike riders similar to those seen in the real task, but without the task environment. During training, they performed 4 rounds with normal vision, and one round with temporal effects and a 3x3 sobel kernel (to accustom the participant to a highly distorted environment). They were given the objective of going from the starting area to a target location across the room. For these training rounds, they were given specific instructions to "run into the person on your right", "run into the trashcan", and "get hit by a biker". The purpose of this was for the subjects to learn the collision system of the experiment. Colliding with a person would result in the person making a

noise, and an on-screen text reading “Collision, back up!”. An arrow was displayed, pointing backwards in the z-direction (see Figure 6.4). The on-screen text would also tell the participant what they ran into, displaying “Person - Walking” or “Fountain”. A subsequent collision was only counted if the participant backed up far enough to remove the notification (0.25 m), and then collided with the same object.

Figure 6.4: To avoid participants walking through virtual objects, they were forced to ’back up’ when a collision was detected. To back up, they had to move 0.25 meters in the negative z-direction. To accomplish this, a UI element would appear informing the participant of the collision. There was text stating “Collision, backup!” along with the name of the object collided with (for example ‘Person, walking’ if they collided with a moving person). Additionally, there was an arrow telling them which direction to back up in. In this figure, the participant (represented by the three colored arrows with blue being forward) has approached the fountain from two separate directions. In the left figure, they approached from the southwest, and were told they needed to go back and to the right to complete their “Back-up”. In the right figure, the participant approached from the east and was told they needed to go directly left to complete their “Back-up”. In both cases, the participant is simply being told to go south (black arrow), back towards the starting location.

After the practice rounds, and at the beginning of each new block, the participant was treated as if they were a prosthesis user and the experimenter was teaching them the environment. They were instructed to look at the arches of the subway station, the fountain by their feet, and the bench to their right with a person on it. When the

participant was ready, they were led to the edge of the bike lane and told this is where they needed to stop before crossing. This was performed with each new block, as each block had a new preprocessing pattern.

At the end of each block, the participant was asked to rate the difficulty of the block on a 1-10 scale.

6.1.4 Data Analysis

Our outcome metrics were trial outcome (successful trial, bike collision, or out of time), time to completion, and number/type of collisions. For each trial, every collision was recorded. This included collisions with moving bikes or the finishing area. We conducted linear mixed-effects analyses using the 'pymer4' package in Python. Our dependent variables were "Success", "Bike Collision" which were modeled as a binomial outcome, and "Total Collisions" modeled as a Poisson outcome as it is non-continuos/count data. We were also interested if the preprocessing method had any effect on the type of collision (e.g. walking people vs stationary objects), and ran additional models with "Stationary Object Collisions", "Walking Person Collisions", and "Stationary People Collisions" modeled as poisson outcomes. We also wanted to know if the perceived difficulty was affected, and modeled "Difficulty" as a gaussian outcome.

To insure a design-driven model Barr et al. (2013), the predictors in the model were "Block Number" (the order they performed the blocks, with the assumption of learning) and "Preprocessing Method". To account for the repeated measures structure of the data, we included a random intercept for SubjectID. Thus the model with "Outcome" being "Successful Finish", "Bike Collision", "Total Collisions", "Stationary Object Collisions", "Walking Person Collisions", "Stationary People Collisions", or

"Difficulty" was:

$$\text{Outcome} \sim C(\text{Block Number}) + C(\text{Preprocessing Method}) + (1|\text{SubjectID})$$

The code, along with all data, is available at
https://github.com/bionicvisionlab/kasowski_2023_smartRaster

6.2 Results

6.2.1 'Smart' Options Result in Less Collisions, but Primarily with Stationary Objects

Both 'smart' options resulted in less total collisions than the naive approach (See Figure 6.5). The mixed linear effects model identified significant differences in "Preprocessing Method" but not "Block Number" for total collisions. When comparing "Naive Edge Detection" with "Smart Edge Detection", participants were estimated to have 0.2 less collisions (95% CI: -0.334, -0.066, $p = 0.003$). When comparing the naive method to "Smart Rasterization", there was an even bigger difference with an average reduction of -0.239 (95% CI: -0.374, -0.104, $p = 0.001$).

We were also curious if the different preprocessing method could have a varying effect depending on the type of obstacle. We separated collisions into "Collisions with Stationary People", "Collisions with Walking People", and "Collisions with Stationary Objects". The linear mixed effects model did not identify any effect of Block Number or Smart Edge Detection, but found that Smart Rasterization led to an estimated 0.878 less collisions (95% CI: -1.755, -0.002, $p = 0.049$) when looking at collisions with walking people.

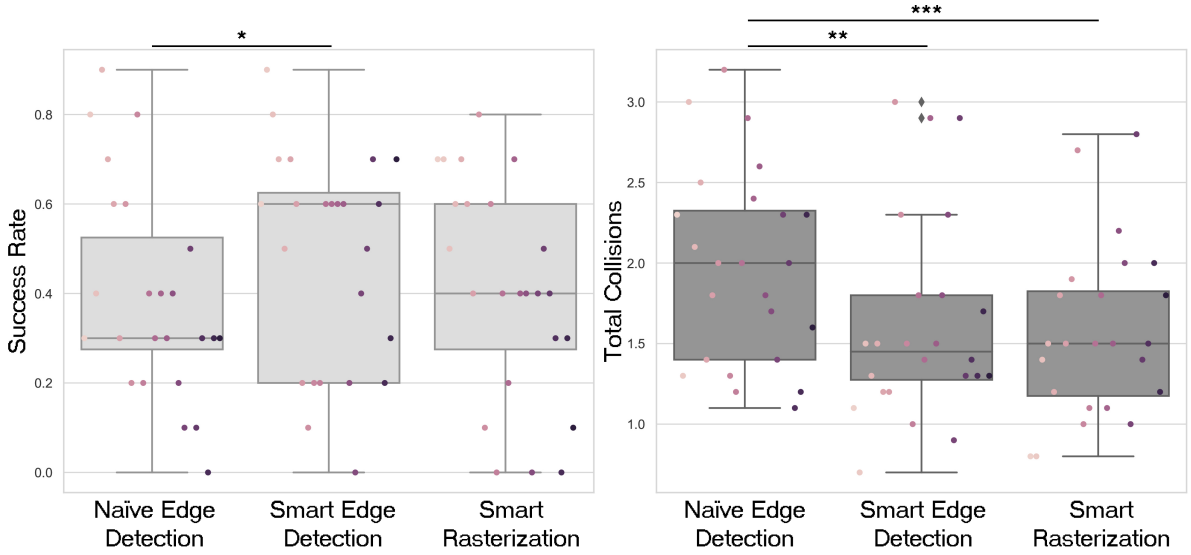


Figure 6.5: Subject averages for 'Success Rate' (Left) and 'Total Collisions' (Right). Significant improvements were identified for the success rate when comparing naive and smart edge detection. Both 'smart' conditions resulted in significantly lower total number of collisions.

Block Number was also not a significant effect in the Stationary Objects model, but both 'smart' methods of preprocessing were found to improve the number of collisions. With 'Smart Edge Detection', participants managed to reduce their collision count by an average of -0.257 (95% CI: -0.458, =0.056, $p = 0.012$), and with 'Smart Rasterization', participants collided an average of 0.308 less collisions (95% CI: -0.512, -0.103, $p = 0.003$).

6.2.2 'Smart' Methods Results in Lower Perceived Difficulty and Higher Success Rate

The mixed linear effects model for "Successful Finish" converged (AIC=919.011), and found significant effects in "Block Number" and "Preprocessing Method". When compared to the first block, the success rate increased by 0.596 (95% CI: 0.198, 0.995, $p < .001$) for the second block, and 0.792 (95% CI: 0.393, 1.192, $p = 0.003$) for the third

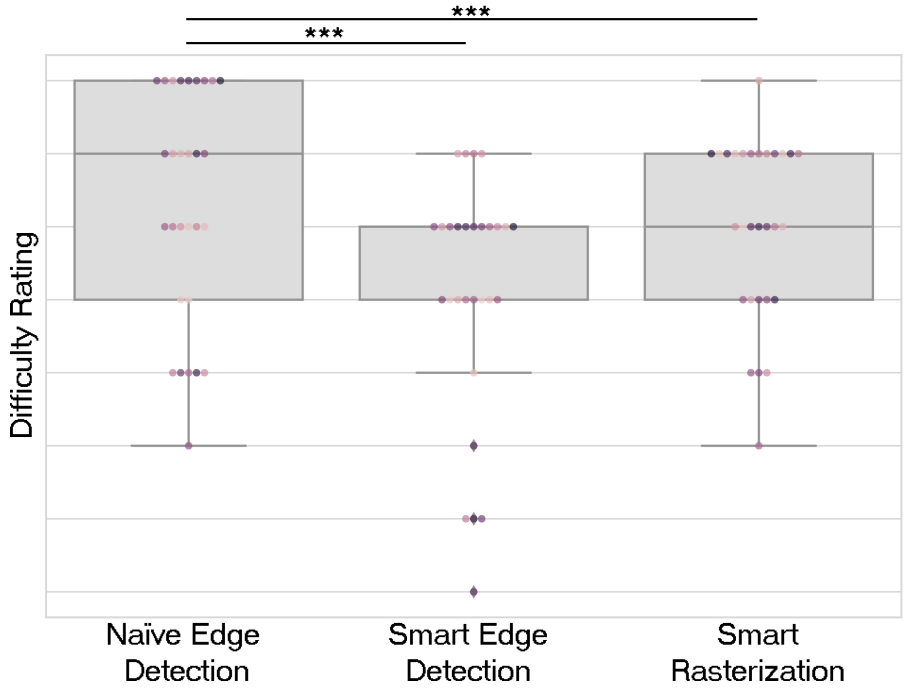


Figure 6.6: Self reported difficulty ratings: At the end of each block, participants were asked to rate the difficulty of the preprocessing method they had just completed. The difficulty scale went from 1 (very easy) to 10 (very hard).

block. This implies a large learning effect with participants performing better on the later blocks. The model found "Smart Edge Detection" to increase the success rate by 0.435 (95% CI: 0.041, 0.829, $p = 0.031$). Surprisingly, while participants did seem to perform better with "Smart Rasterization" (see Fig. 6.2), this difference failed to reach significance.

The mixed linear effects model for "Difficulty" converged (AIC=2092.259), and found significant effects in "Block Number" and "Preprocessing Method" (see Figure 6.6). When comparing the second block to the first block completed, participants tended to change their difficulty rating by -0.673 (95% CI: -0.845, -0.501, $p < 0.001$). When comparing the first and third block, the difference was even greater with a change of -0.964 (95% CI: -1.137, -0.792, $p < 0.001$). Both 'smart' preprocessing methods were also rated significantly less difficult than the naive approach. When using "Smart Edge

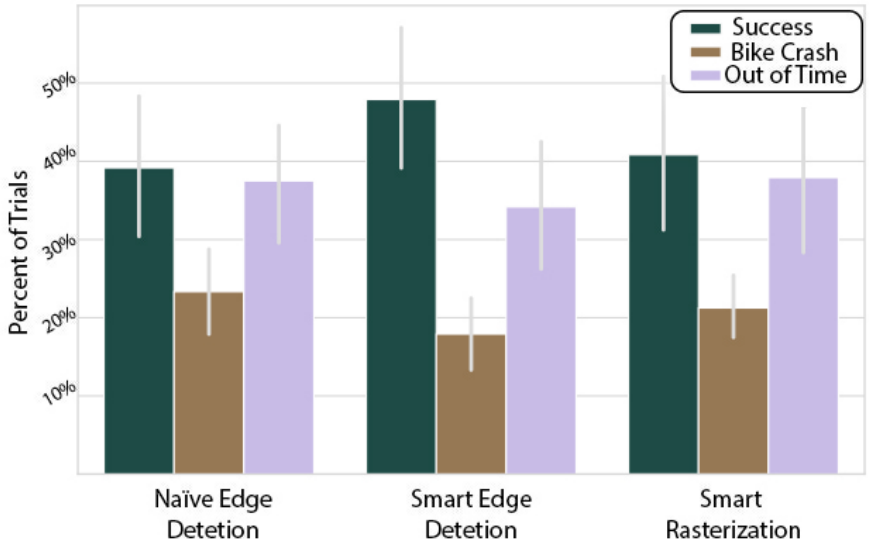


Figure 6.7: Trial completion outcome percentages for the different processing methods. In all three scenarios, participants were able to complete the objective in the majority of trials. A trial was considered a success if the participant was able to navigate inside the subway station and to the designated side (left/right). Trials could also end when the user ran out of time or collided with a moving cyclist.

Detection", participants changed their difficulty ratings by an average of -1.083 (95% CI: -1.255, -0.911, $p < 0.001$), and by an average of -0.411 (95% CI: -0.583, -0.239, $p < 0.001$) when they were using "Smart Rasterization".

6.2.3 Neither 'Smart' Condition Improved Bike Path Safety

Our original objective was finding a way to utilize the temporal dynamics of electrode rasterization to elicit more information and increase user safety. While participants were able to reach their destination more often, perceived less difficulty, and had overall less collisions, there was no significant effect on the number of bike collisions (see Figure 6.7).

6.3 Discussion

This study has demonstrated smart preprocessing has the capacity to improve collision rate, help the user successfully reach their target, and has a lower perceived difficulty to the user. This has immediate implications for current device users, but our main hypothesis was 'Smart Rasterization' would not only help avoid collisions, but would help the user reach the target location more often and more safely. Our results failed to demonstrate any significant difference in the number of bike collisions between the different preprocessing methods, and this could be for a number of reasons. After each subject was finished, they were asked if they implemented any strategies or had any thoughts about the simulation. While responses were varied from "that was a lot of fun" to "that was extremely frustrating", one response that came up repeatedly was that participants would stop at the edge of the bike path and wait until they didn't hear any bikes nearby. The experiment utilized spatial audio, and bike sounds were attached to the physical location of the bikes. In retrospect, this is not surprising as this is most likely the same strategy someone with ultra low vision would implement. Additionally, while a large variety of timings were piloted, only one smart rasterization pattern was fully examined in this experiment. It is possible that different timings, different segmentation groups, or a different order within the groups may result in better performance. Another future avenue of exploration would be allowing the participant to switch between the various groupings through a controller. This would likely take a large degree of training to be successful, but it does provide the user with more control over their perception of the environment. In summary, smart preprocessing techniques have been demonstrated to improve performance on certain metrics, but more work is needed to definitively conclude that temporal segmentation provides additional benefits to static segmentation.

Chapter 7

Summary and Conclusions

Visual prostheses have great potential, but current devices have been unable to meet consumer expectations and have failed to be successful commercially. A major challenge is fully understanding what patients see with the various devices. Figuring this out requires collaborative work from multiple fields, including neuroscientists, vision researchers, computer scientists, and researchers properly trained in psychophysics research. In Chapter 3, I introduced software that enabled the practical feasibility of simulating prosthetic vision under realistic conditions; BionicVisionXR incorporates a neurophysiologically inspired model and is built on clinical reports of spatiotemporal distortions in device users. It continues to go well beyond any other current models of prosthetic vision by accounting for gaze congruency, and by considering practical/engineering limitations on the number of electrodes you can stimulate simultaneously.

In Chapter 4, I provided evidence that these things do matter, and that by ignoring spatial distortions or immersion in previous work, simulation studies likely misrepresented the experience of prosthesis users. I was able to show that immersion is extremely important for a task requiring proprioception, but may not be as important if

the researcher is performing simple two dimensional recognition tasks. It is also worth noting that performance on these tasks still had significant room for improvement, indicating there is still a lot to learn about stimulus optimization and device designs.
In Chapter 5, I discussed the first study of its kind, investigating electrode rasterization with an XR model of prosthetic vision. I tested common rasterization strategies for suspected biases and found that apparent motion from certain rasterization patterns does bias motion discrimination in the expected direction. I also identified an easily implementable rasterization strategy that should serve as a 'general purpose' rasterization pattern, potentially improving performance across a variety of tasks.
In Chapter 6, I introduced smart raster patterns which could be implemented with computer vision on a near-future device. This was the first study to attempt to introduce additional information by using temporal segmentation of visually segmented groups. While the certain smart rastering strategy used did not significantly improve performance, simplifying the scene through 'Smart Edge Detection' was able to improve success rate. Additionally, both smart techniques did significantly lower the overall number of collisions, highlighting the value of computer vision and potentially other forms of artificial intelligence for improving prosthetic vision.
Overall, the contributions of this thesis span multiple disciplines including computational neuroscience, computer science/engineering, and psychology. The present work introduces simple mathematical models for the temporal dynamics of neuron desensitization, and demonstrates how neurophysical models can be used not only in general research, but in practical applications like prosthetics. The work leaves behind open source material that could be of interest to neuroscientists building novel models, computer scientists working in XR, and those in the psychological and brain sciences interested in visual sciences or navigation. Like any other researcher who respects their field, I sincerely hope my work has contributed to our societal knowledge

and has laid the foundation for future discoveries to build upon.

Appendix A

Eye tracking accuracy of the HTC Vive Pro

To measure the precision of the VIVE's built-in eye tracker, we had participants move their eyes to track a moving, on-screen dot. This was performed using a custom script that moved the dot (sized at $\sim 2.4^o$ visual angle) randomly between the 4 corners located halfway between the center and edges of the screen. The dot moved ten times, traversing the distance between points in 2.5 ± 0.5seconds and staying at each location for 1.5 seconds. The angular distance between the center of the dot and the user's gaze location was measured every 0.1 s. Measurements were taken during fixation (when the dot was stationary) and during pursuit. The angular error was $(1.904 \pm 2.048)^o$ during fixations and $(1.838 \pm 1.660)^o$ during pursuit. The two distributions were not significantly different (t-test for non-equal variances, $p > 0.27$). Overall, 94.1% of measurements had an angular error less than 5^o, and 80 % had angular errors less than 3^o (see Figure A.1).

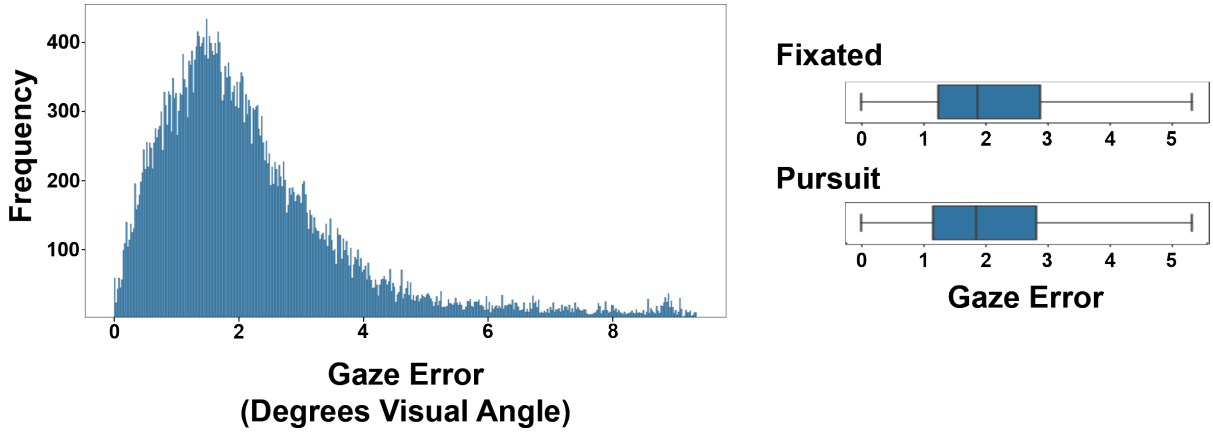


Figure A.1: The histogram (Left) contains all data points below four standard deviations from the mean. The distribution plots (right) show gaze error (outliers removed) when the dot is stationary (fixated) or moving (pursuit).

Appendix B

Checkerboard Rasterization

The checkerboard rasterization pattern was designed to be expandable to any square shaped array. However, to avoid apparent motion the groups must be presented in a non linear sequence (see Figure B.1). It creates the groups by 'knowing' how many electrodes should be in each row, and doing row-wise processing. For each row, it loops through the electrodes and determines which should be in the current group. For example, if there were 10 electrodes and 2 should be active in the first row. In group '0', it would decide electrode 0 and 5 should be active. For odd rows, it shifts these values by half the size of the raster group, effectively maximizing the space between any adjacent electrodes.

```
for (int i = 0; i < rasterizeGroups; i++) {
    int currElectrode = 0;
    Electrode[] electrodes = bs.electrodes;
    Electrode[] subset = new Electrode[electrodes.Length / rasterizeGroups];
    int row = 0;
    int count = 0;

    while (row < vm.numberYelectrodes) {
        for(int j=0; j<(vm.numberXelectrodes / rasterizeGroups); j++) {
            currElectrode = (row * vm.numberXelectrodes) + i + (rasterizeGroups * j);

            if (row % 2 == 0) {
                subset[count] = electrodes[currElectrode];
                count += 1; }
            else{

                currElectrode = currElectrode + (int) Math.Ceiling(rasterizeGroups / 2.0);

                currElectrode = currElectrode >= ((row+1)* vm.numberXelectrodes)?
                     currElectrode - vm.numberXelectrodes : currElectrode;

                subset[count] = electrodes[currElectrode];
                count += 1; }
        }
        row++; }
    int[] checkerboardOrder = {0, 4, 2, 1, 3 };
    rasterizedGroups[checkerboardOrder[i]] = (Electrode[]) subset.Clone(); } } }
```

Figure B.1:

Bibliography

Al-Atabany, W. I., Tong, T., and Degenaar, P. A. (2010). Improved content aware scene retargeting for retinitis pigmentosa patients. *BioMedical Engineering OnLine*, 9(1):52.

Avraham, D., Jung, J.-H., Yitzhaky, Y., and Peli, E. (2021). Retinal prosthetic vision simulation: temporal aspects. *Journal of Neural Engineering*, 18(4).

Ayton, L. N., Barnes, N., Dagnelie, G., Fujikado, T., Goetz, G., Hornig, R., Jones, B. W., Muqit, M. M. K., Rathbun, D. L., Stingl, K., Weiland, J. D., and Petoe, M. A. (2020). An update on retinal prostheses. *Clinical Neurophysiology*, 131(6):1383–1398.

Ayton, L. N., Blamey, P. J., Guymer, R. H., Luu, C. D., Nayagam, D. A. X., Sinclair, N. C., Shivdasani, M. N., Yeoh, J., McCombe, M. F., Briggs, R. J., Opie, N. L., Villalobos, J., Dimitrov, P. N., Varsamidis, M., Petoe, M. A., McCarthy, C. D., Walker, J. G., Barnes, N., Burkitt, A. N., Williams, C. E., Shepherd, R. K., Allen, P. J., and Consortium, f. t. B. V. A. R. (2014). First-in-Human Trial of a Novel Suprachoroidal Retinal Prosthesis. *PLOS ONE*, 9(12):e115239.

Bao, M. and Engel, S. A. (2019). Augmented Reality as a Tool for Studying Visual Plasticity: 2009 to 2018. *Current Directions in Psychological Science*, 28(6):574–580.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

Barry, M. P., Armenta Salas, M., Patel, U., Wuyyuru, V., Niketeghad, S., Bosking, W. H., Yoshor, D., Dorn, J. D., and Pouratian, N. (2020). Video-mode percepts are smaller than sums of single-electrode phosphenes with the Orion® visual cortical prosthesis. *Investigative Ophthalmology & Visual Science*, 61(7):927.

Barry, M. P. and Dagnelie, G. (2016). Hand-Camera Coordination Varies over Time in Users of the Argus® II Retinal Prosthesis System. *Frontiers in Systems Neuroscience*, 10.

Beauchamp, M. S., Oswalt, D., Sun, P., Foster, B. L., Magnotti, J. F., Niketeghad, S., Pouratian, N., Bosking, W. H., and Yoshor, D. (2020). Dynamic Stimulation of Visual Cortex Produces Form Vision in Sighted and Blind Humans. *Cell*, 181(4):774–783.e5.

Behrend, M. R., Ahuja, A. K., Humayun, M. S., Chow, R. H., and Weiland, J. D. (2011). Resolution of the Epiretinal Prosthesis is not Limited by Electrode Size. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(4):436–442.

Beyeler, M., Boynton, G. M., Fine, I., and Rokem, A. (2019a). Model-Based Recommendations for Optimal Surgical Placement of Epiretinal Implants. In Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., Zhou, S., Yap, P.-T., and Khan, A., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Lecture Notes in Computer Science, pages 394–402. Springer International Publishing.

Beyeler, M., Nanduri, D., Weiland, J. D., Rokem, A., Boynton, G. M., and Fine, I. (2019b). A model of ganglion axon pathways accounts for percepts elicited by retinal implants. *Scientific Reports*, 9(1):9199. Number: 1 Publisher: Nature Publishing Group.

Beyeler, M., Rokem, A., Boynton, G. M., and Fine, I. (2017). Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies. *Journal of Neural Engineering*, 14(5):051003. Publisher: IOP Publishing.

Beyeler, M. and Sanchez-Garcia, M. (2022). Towards a Smart Bionic Eye: AI-powered artificial vision for the treatment of incurable blindness. *Journal of Neural Engineering*, 19(6):063001. Publisher: IOP Publishing.

Bosking, W. H., Sun, P., Ozker, M., Pei, X., Foster, B. L., Beauchamp, M. S., and Yoshor, D. (2017). Saturation in Phosphene Size with Increasing Current Levels Delivered to Human Visual Cortex. *Journal of Neuroscience*, 37(30):7188–7197. Publisher: Society for Neuroscience Section: Research Articles.

Bourkiza, B., Vurro, M., Jeffries, A., and Pezaris, J. S. (2013). Visual Acuity of Simulated Thalamic Visual Prostheses in Normally Sighted Humans. *PLOS ONE*, 8(9):e73592. Publisher: Public Library of Science.

Bourne, R. R. A., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., Leasher, J., Limburg, H., Naidoo, K., Pesudovs, K., Resnikoff, S., Silvester, A., Stevens, G. A., Tahhan, N., Wong, T. Y., Taylor, H. R., Bourne, R., Ackland, P., Arditi, A., Barkana, Y., Bozkurt, B., Braithwaite, T., Bron, A., Budenz, D., Cai, F., Casson, R., Chakravarthy, U., Choi, J., Cicinelli, M. V., Congdon, N., Dana, R., Dandona, R., Dandona, L., Das, A., Dekaris, I., Monte, M. D., Deva, J., Dreer, L., Ellwein, L., Frazier, M., Frick, K., Friedman, D., Furtado, J., Gao, H., Gazzard, G., George, R., Gichuhi, S., Gonzalez, V., Hammond, B., Hartnett, M. E., He, M., Hejtmancik, J., Hirai, F., Huang, J., Ingram, A., Javitt, J., Jonas, J., Joslin, C., Keeffe, J., Kempen, J., Khairallah, M., Khanna, R., Kim, J., Lambrou, G., Lansingh, V. C., Lanzetta, P., Leasher, J., Lim, J., Limburg, H., Mansouri, K., Mathew, A., Morse, A., Munoz, B., Musch, D., Naidoo, K., Nangia, V., Palaiou, M., Parodi, M. B., Pena, F. Y., Pesudovs, K., Peto, T., Quigley, H., Raju, M., Ramulu, P., Resnikoff,

S., Robin, A., Rossetti, L., Saaddine, J., Sandar, M., Serle, J., Shen, T., Shetty, R., Sieving, P., Silva, J. C., Silvester, A., Sitorus, R. S., Stambolian, D., Stevens, G., Taylor, H., Tejedor, J., Tielsch, J., Tsilimbaris, M., Meurs, J. v., Varma, R., Virgili, G., Volmink, J., Wang, Y. X., Wang, N.-L., West, S., Wiedemann, P., Wong, T., Wormald, R., and Zheng, Y. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, 5(9):e888–e897. Publisher: Elsevier.

Boyle, J. R., Maeder, A. J., and Boles, W. W. (2008). Region-of-interest processing for electronic visual prostheses. *Journal of Electronic Imaging*, 17(1):013002. Publisher: International Society for Optics and Photonics.

Cao, X., Li, H., Lu, Z., Chai, X., and Wang, J. (2017). Eye-hand coordination using two irregular phosphene maps in simulated prosthetic vision for retinal prostheses. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–5.

Caspi, A., Barry, M. P., Patel, U. K., Salas, M. A., Dorn, J. D., Roy, A., Niketeghad, S., Greenberg, R. J., and Pouratian, N. (2021). Eye movements and the perceived location of phosphenes generated by intracranial primary visual cortex stimulation in the blind. *Brain Stimulation*, 14(4):851–860.

Caspi, A., Roy, A., Wuyyuru, V., Rosendall, P. E., Harper, J. W., Katyal, K. D., Barry, M. P., Dagnelie, G., and Greenberg, R. J. (2018). Eye Movement Control in the Argus II Retinal-Prosthesis Enables Reduced Head Movement and Better Localization Precision. *Investigative Ophthalmology & Visual Science*, 59(2):792–802.

Caspi, A. and Zivotofsky, A. Z. (2015). Assessing the utility of visual acuity measures in visual prostheses. *Vision Research*, 108:77–84.

Chang, M., Kim, H. S., Shin, J. H., and Park, K.-S. (2012). Facial identification in very low-resolution images simulating prosthetic vision. *Journal of Neural Engineering*, 9(4):046012.

Chen, S. C., Suaning, G. J., Morley, J. W., and Lovell, N. H. (2009). Simulating prosthetic vision: I. Visual models of phosphenes. *Vision Research*, 49(12):1493–1506.

Christie, B., Sadeghi, R., Kartha, A., Caspi, A., Tenore, F. V., Klatzky, R. L., Dagnelie, G., and Billings, S. (2022). Sequential epiretinal stimulation improves discrimination in simple shape discrimination tasks only. *Journal of Neural Engineering*, 19(3):036033. Publisher: IOP Publishing.

Cruz, L. d., Coley, B. F., Dorn, J., Merlini, F., Filley, E., Christopher, P., Chen, F. K., Wuyyuru, V., Sahel, J., Stanga, P., Humayun, M., Greenberg, R. J., Dagnelie, G., and Group, f. t. A. I. S. (2013). The Argus II epiretinal prosthesis system allows

letter and word reading and long-term function in patients with profound vision loss. *British Journal of Ophthalmology*, 97(5):632–636. Publisher: BMJ Publishing Group Ltd Section: Clinical science.

Dagnelie, G., Keane, P., Narla, V., Yang, L., Weiland, J., and Humayun, M. (2007). Real and virtual mobility performance in simulated prosthetic vision. *J Neural Eng*, 4(1):S92–101.

Denis, G., Jouffrais, C., Vergnieux, V., and Macé, M. (2013). Human faces detection and localization with simulated prosthetic vision. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 61–66.

Dobelle, W. H. (2000). Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex. *ASAIO Journal*, 46(1):3.

Dobelle, W. H. and Mladejovsky, M. G. (1974). Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind. *The Journal of Physiology*, 243(2):553–576. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1974.sp010766.

Elkin, L. A., Kay, M., Higgins, J. J., and Wobbrock, J. O. (2021). An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 754–768. Association for Computing Machinery, New York, NY, USA.

Erickson-Davis, C. and Korzybska, H. (2021). What do blind people "see" with retinal prostheses? Observations and qualitative reports of epiretinal implant users. *PLOS ONE*, 16(2):e0229189.

Evans, J. R., Gordon, J., Abramov, I., Mladejovsky, M. G., and Dobelle, W. H. (1979). Brightness of phosphenes elicited by electrical stimulation of human visual cortex. *Sensory Processes*, 3(1):82–94. Place: US Publisher: Academic Press.

Ferlauto, L., Airaghi Leccardi, M. J. I., Chenais, N. A. L., Gilliéron, S. C. A., Vagni, P., Bevilacqua, M., Wolfensberger, T. J., Sivula, K., and Ghezzi, D. (2018). Design and validation of a foldable and photovoltaic wide-field epiretinal prosthesis. *Nature Communications*, 9(1):992. Number: 1 Publisher: Nature Publishing Group.

Fernandez, E. (2018). Development of visual Neuroprostheses: trends and challenges. *Bioelectronic Medicine*, 4(1):12.

Fernández, E., Alfaro, A., Soto-Sánchez, C., Gonzalez-Lopez, P., Lozano, A. M., Peña, S., Grima, M. D., Rodil, A., Gómez, B., Chen, X., Roelfsema, P. R., Rolston, J. D., Davis, T. S., and Normann, R. A. (2021). Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex. *The Journal of Clinical Investigation*, 131(23). Publisher: American Society for Clinical Investigation.

Fine, I. and Boynton, G. M. (2015). Pulse trains to percepts: the challenge of creating a perceptually intelligible world with sight recovery technologies. *Philos Trans R Soc Lond B Biol Sci*, 370(1677):20140208.

Fornos, A. P., Sommerhalder, J., Cruz, L. d., Sahel, J. A., Mohand-Said, S., Hafezi, F., and Pelizzone, M. (2012). Temporal Properties of Visual Perception on Electrical Stimulation of the Retina. *Investigative Ophthalmology & Visual Science*, 53(6):2720–2731. Publisher: The Association for Research in Vision and Ophthalmology.

Fornos, A. P., Sommerhalder, J., and Pelizzone, M. (2011). Reading with a simulated 60-channel implant. *Frontiers in Neuroscience*, 5:57.

Han, N., Srivastava, S., Xu, A., Klein, D., and Beyeler, M. (2021). Deep Learning–Based Scene Simplification for Bionic Vision. In *Augmented Humans Conference 2021*, pages 45–54, Rovaniemi Finland. ACM.

He, Y., Sun, S. Y., Roy, A., Caspi, A., and Montezuma, S. R. (2020). Improved mobility performance with an artificial vision therapy system using a thermal sensor. *Journal of Neural Engineering*, 17(4):045011.

Horne, L., Alvarez, J., McCarthy, C., Salzmann, M., and Barnes, N. (2016). Semantic labeling for prosthetic vision. *Computer Vision and Image Understanding*, 149:113–125.

Horsager, A., Greenberg, R. J., and Fine, I. (2010). Spatiotemporal Interactions in Retinal Prosthesis Subjects. *Investigative Ophthalmology & Visual Science*, 51(2):1223–1233.

Horsager, A., Greenwald, S. H., Weiland, J. D., Humayun, M. S., Greenberg, R. J., McMahon, M. J., Boynton, G. M., and Fine, I. (2009). Predicting Visual Sensitivity in Retinal Prosthesis Patients. *Investigative Ophthalmology & Visual Science*, 50(4):1483–1491.

Jansonius, N. M., Schiefer, J., Nevalainen, J., Paetzold, J., and Schiefer, U. (2012). A mathematical model for describing the retinal nerve fiber bundle trajectories in the human eye: Average course, variability, and influence of refraction, optic disc size and optic disc position. *Experimental Eye Research*, 105:70–78.

Jiang, H., Li, H., Liang, J., and Chai, X. (2020). A hierarchical image processing strategy for artificial retinal prostheses. In *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, pages 359–362.

Josh, H., Mann, C., Kleeman, L., and Lui, W. L. D. (2013). Psychophysics testing of bionic vision image processing algorithms using an FPGA Hatpack. In *2013 IEEE International Conference on Image Processing*, pages 1550–1554.

Kardong-Edgren, S. S., Farra, S. L., Alinier, G., and Young, H. M. (2019). A Call to Unify Definitions of Virtual Reality. *Clinical Simulation in Nursing*, 31:28–34.

Kasowski, J. and Beyeler, M. (2022). Immersive Virtual Reality Simulations of Bionic Vision. In *Augmented Humans 2022*, AHs 2022, pages 82–93, New York, NY, USA. Association for Computing Machinery.

Kasowski, J., Johnson, B. A., Neydavood, R., Akkaraju, A., and Beyeler, M. (2023). A systematic review of extended reality (XR) for understanding and augmenting vision loss. *Journal of Vision*, 23(5):5.

Kasowski, J., Wu, N., and Beyeler, M. (2021). Towards Immersive Virtual Reality Simulations of Bionic Vision. In *Augmented Humans Conference 2021*, pages 313–315, Rovaniemi Finland. ACM.

Li, H., Han, T., Wang, J., Lu, Z., Cao, X., Chen, Y., Li, L., Zhou, C., and Chai, X. (2017). A real-time image optimization strategy based on global saliency detection for artificial retinal prostheses. *Information Sciences*, 415-416:1–18.

Li, H., Su, X., Wang, J., Kan, H., Han, T., Zeng, Y., and Chai, X. (2018). Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision. *Artificial Intelligence in Medicine*, 84:64–78.

Lorach, H., Goetz, G., Smith, R., Lei, X., Mandel, Y., Kamins, T., Mathieson, K., Huie, P., Harris, J., Sher, A., and Palanker, D. (2015). Photovoltaic restoration of sight with high visual acuity. *Nat Med*, 21(5):476–82.

Lu, Y., Chen, P., Zhao, Y., Shi, J., Ren, Q., and Chai, X. (2012). Estimation of simulated phosphene size based on tactile perception. *Artificial Organs*, 36(1):115–120.

Lui, W. L. D., Browne, D., Kleeman, L., Drummond, T., and Li, W. H. (2011). Transformative reality: Augmented reality for visual prostheses. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 253–254.

Luo, Y. H. and da Cruz, L. (2016a). The Argus((R)) II Retinal Prosthesis System. *Prog Retin Eye Res*, 50:89–107.

Luo, Y. H.-L. and da Cruz, L. (2016b). The Argus® II Retinal Prosthesis System. *Progress in Retinal and Eye Research*, 50:89–107.

Luo, Y. H.-L., Zhong, J. J., Clemo, M., and da Cruz, L. (2016). Long-term Repeatability and Reproducibility of Phosphene Characteristics in Chronically Implanted Argus II Retinal Prosthesis Subjects. *American Journal of Ophthalmology*, 170:100–109.

Macé, M. J.-M., Guivarch, V., Denis, G., and Jouffrais, C. (2015). Simulated Prosthetic Vision: The Benefits of Computer-Based Object Recognition and Localization. *Artificial Organs*, 39(7):E102–113.

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, 6.

McCarthy, C., Walker, J. G., Lieby, P., Scott, A., and Barnes, N. (2015). Mobility and low contrast trip hazard avoidance using augmented depth. *Journal of Neural Engineering*, 12(1):016003.

Miller, H. L. and Bugnariu, N. L. (2016). Level of Immersion in Virtual Environments Impacts the Ability to Assess and Teach Social Skills in Autism Spectrum Disorder. *Cyberpsychology, Behavior, and Social Networking*, 19(4):246–256.

Oswalt, D., Bosking, W., Sun, P., Sheth, S. A., Niketeghad, S., Salas, M. A., Patel, U., Greenberg, R., Dorn, J., Pouratian, N., Beauchamp, M., and Yoshor, D. (2021). Multi-electrode stimulation evokes consistent spatial patterns of phosphenes and improves phosphene mapping in blind subjects. *Brain Stimulation*, 14(5):1356–1372.

Paraskevoudi, N. and Pezaris, J. S. (2019). Eye Movement Compensation and Spatial Updating in Visual Prosthetics: Mechanisms, Limitations and Future Directions. *Frontiers in Systems Neuroscience*, 12.

Paraskevoudi, N. and Pezaris, J. S. (2021). Full gaze contingency provides better reading performance than head steering alone in a simulation of prosthetic vision. *Scientific Reports*, 11(1):11121.

Pasch, M., Bianchi-Berthouze, N., van Dijk, B., and Nijholt, A. (2009). Immersion in Movement-Based Interaction. In Nijholt, A., Reidsma, D., and Hondorp, H., editors, *Intelligent Technologies for Interactive Entertainment*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pages 169–180, Berlin, Heidelberg. Springer.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Perez-Yus, A., Bermudez-Cameo, J., Guerrero, J. J., and Lopez-Nicolas, G. (2017). Depth and Motion Cues with Phosphene Patterns for Prosthetic Vision. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 1516–1525.

Rassia, K. E. K. and Pezaris, J. S. (2018). Improvement in reading performance through training with simulated thalamic visual prostheses. *Scientific Reports*, 8(1):16310. Number: 1 Publisher: Nature Publishing Group.

Rizzo, J. F., Wyatt, J., Loewenstein, J., Kelly, S., and Shire, D. (2003). Perceptual Efficacy of Electrical Stimulation of Human Retina with a Microelectrode Array during Short-Term Surgical Trials. *Investigative Ophthalmology & Visual Science*, 44(12):5362–5369.

Rizzo, S., Belting, C., Cinelli, L., Allegrini, L., Genovesi-Ebert, F., Barca, F., and di Bartolo, E. (2014). The Argus II Retinal Prosthesis: 12-Month Outcomes from a Single-Study Center. *American Journal of Ophthalmology*, 157(6):1282–1290.

Sabbah, N., Authié, C. N., Sanda, N., Mohand-Said, S., Sahel, J.-A., and Safran, A. B. (2014). Importance of Eye Position on Spatial Localization in Blind Subjects Wearing an Argus II Retinal Prosthesis. *Investigative Ophthalmology & Visual Science*, 55(12):8259–8266.

Sadeghi, R., Kartha, A., Barry, M. P., Bradley, C., Gibson, P., Caspi, A., Roy, A., and Dagnelie, G. (2021). Glow in the dark: Using a heat-sensitive camera for blind individuals with prosthetic vision. *Vision Research*, 184:23–29.

Sanchez-Garcia, M., Martinez-Cantin, R., Bermudez-Cameo, J., and Guerrero, J. J. (2020). Influence of field of view in visual prostheses design: Analysis with a VR system. *Journal of Neural Engineering*, 17(5):056002.

Sanchez-Garcia, M., Martinez-Cantin, R., and Guerrero, J. J. (2019). Indoor Scenes Understanding for Visual Prosthesis with Fully Convolutional Networks. In *VISIGRAPP*.

Schmidt, E. M., Bak, M. J., Hambrecht, F. T., Kufta, C. V., O'Rourke, D. K., and Vallabhanath, P. (1996). Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex. *Brain: A Journal of Neurology*, 119 (Pt 2):507–522.

Sight, S. (2013). *Argus® II Retinal Prosthesis System Surgeon Manual*. Number 900029-001 Rev C. Second Sight Medical Products, Inc., Sylmar, CA.

Stacey, A., Li, Y., and Barnes, N. (2011). A salient information processing system for bionic eye with application to obstacle avoidance. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, volume 2011, pages 5116–5119.

Stingl, K., Bartz-Schmidt, K. U., Besch, D., Braun, A., Bruckmann, A., Gekeler, F., Greppmaier, U., Hipp, S., Hortdorfer, G., Kernstock, C., Koitschev, A., Kusnyerik, A., Sachs, H., Schatz, A., Stingl, K. T., Peters, T., Wilhelm, B., and Zrenner, E. (2013). Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS. *Proc Biol Sci*, 280(1757):20130077.

Thorn, J. T., Chenais, N. A. L., Hinrichs, S., Chatelain, M., and Ghezzi, D. (2022). Virtual reality validation of naturalistic modulation strategies to counteract fading in retinal stimulation. *Journal of Neural Engineering*, 19(2):026016. Publisher: IOP Publishing.

Thorn, J. T., Migliorini, E., and Ghezzi, D. (2020). Virtual reality simulation of epiretinal stimulation highlights the relevance of the visual angle in prosthetic vision. *Journal of Neural Engineering*, 17(5):056019. Publisher: IOP Publishing.

Titchener, S. A., Shivdasani, M. N., Fallon, J. B., and Petoe, M. A. (2018). Gaze Compensation as a Technique for Improving Hand-Eye Coordination in Prosthetic Vision. *Translational Vision Science & Technology*, 7(1):2.

Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, 5(2):99–114. Publisher: [Wiley, International Biometric Society].

Vergnieux, V., Macé, M. J.-M., and Jouffrais, C. (2017). Simplification of Visual Rendering in Simulated Prosthetic Vision Facilitates Navigation. *Artificial Organs*, 41(9):852–861.

Vurro, M., Crowell, A. M., and Pezaris, J. S. (2014). Simulation of thalamic prosthetic vision: reading accuracy, speed, and acuity in sighted humans. *Frontiers in Human Neuroscience*, 8.

Wang, J., Li, H., Fu, W., Chen, Y., Li, L., Lyu, Q., Han, T., and Chai, X. (2016). Image Processing Strategies Based on a Visual Saliency Model for Object Recognition Under Simulated Prosthetic Vision. *Artificial Organs*, 40(1):94–100.

Wang, J., Wu, X., Lu, Y., Wu, H., Kan, H., and Chai, X. (2014). Face recognition in simulated prosthetic vision: face detection-based image processing strategies. *Journal of Neural Engineering*, 11(4):046009.

Wang, L., Sharifian, F., Napp, J., Nath, C., and Pollmann, S. (2018). Cross-task perceptual learning of object recognition in simulated retinal implant perception. *Journal of Vision*, 18(13):22–22. Publisher: The Association for Research in Vision and Ophthalmology.

Weiland, J. D., Parikh, N., Pradeep, V., and Medioni, G. (2012). Smart image processing system for retinal prosthesis. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, volume 2012, pages 300–303.

Wilke, R. G. H., Moghadam, G. K., Lovell, N. H., Suaning, G. J., and Dokos, S. (2011). Electric crosstalk impairs spatial resolution of multi-electrode arrays in retinal implants. *Journal of Neural Engineering*, 8(4):046016.

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 143–146, Vancouver BC Canada. ACM.

Wu, H., Wang, J., Li, H., and Chai, X. (2014). Prosthetic vision simulating system and its application based on retinal prosthesis. In *2014 International Conference on Information Science, Electronics and Electrical Engineering*, volume 1, pages 425–429.

Ying, Z., Xiulin, G., Qi, L., and Guangqi, J. (2018). Recognition of virtual maze scene under simulated prosthetic vision. In *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*, pages 1–5.

Yücel, E. I., Sadeghi, R., Kartha, A., Montezuma, S. R., Dagnelie, G., Rokem, A., Boynton, G. M., Fine, I., and Beyeler, M. (2022). Factors affecting two-point discrimination in Argus II patients. *Frontiers in Neuroscience*, 16:901337.

Zhao, Y., Lu, Y., Tian, Y., Li, L., Ren, Q., and Chai, X. (2010). Image processing based recognition of images with a limited number of pixels using simulated prosthetic vision (51853). *Information Sciences*.

Zhao, Y., Lu, Y., Zhao, J., Wang, K., Ren, Q., Wu, K., and Chai, X. (2011). Reading pixelized paragraphs of Chinese characters using simulated prosthetic vision. *Investigative Ophthalmology & Visual Science*, 52(8):5987–5994.