Immersive Virtual Reality Simulations of Bionic Vision

Justin Kasowski
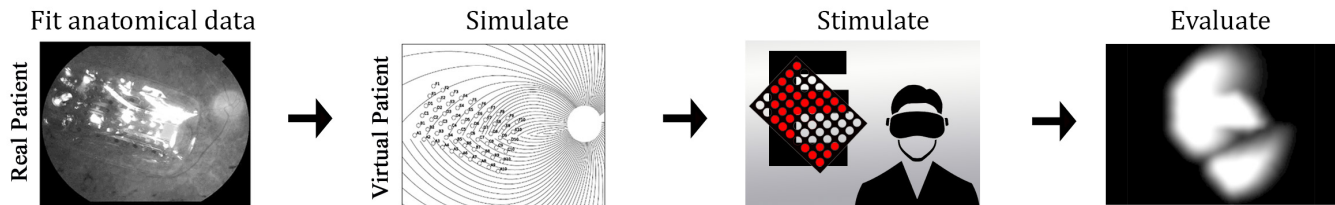University of California, Santa Barbara
Santa Barbara, CA, USA
justin_kasowski@ucsb.edu

Michael Beyeler
University of California, Santa Barbara
Santa Barbara, CA, USA
mbeyeler@ucsb.edu

Figure 1: Immersive virtual reality simulations of bionic vision. A microelectrode array is implanted in the eye to stimulate the retina → Anatomical data is used to position a simulated electrode array on a simulated retina to create a "virtual patient" → Visual input from a virtual reality environment acts as stimulus for the simulated implant to generate a realistic prediction of simulated prosthetic vision (SPV) → The rendered SPV image is presented to the virtual patient and behavioral metrics are recorded

ABSTRACT

Bionic vision uses neuroprostheses to restore useful vision to people living with incurable blindness. However, a major outstanding challenge is predicting what people "see" when they use their devices. The limited field of view of current devices necessitates head movements to scan the scene, which is difficult to simulate on a computer screen. In addition, many computational models of bionic vision lack biological realism. To address these challenges, we present VR-SPV, an open-source virtual reality toolbox for simulated prosthetic vision that uses a psychophysically validated computational model to allow sighted participants to "see through the eyes" of a bionic eye user. To demonstrate its utility, we systematically evaluated how clinically reported visual distortions affect performance in a letter recognition and an immersive obstacle avoidance task. Our results highlight the importance of using an appropriate phosphene model when predicting visual outcomes for bionic vision.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility technologies**; **Virtual reality**; *Empirical studies in visualization.*

KEYWORDS

retinal implant, virtual reality, simulated prosthetic vision

ACM Reference Format:
Justin Kasowski and Michael Beyeler. 2022. Immersive Virtual Reality Simulations of Bionic Vision. In *Augmented Humans 2022 (AHs 2022), March*

1 INTRODUCTION

The World Health Organization has estimated that by the year 2050, roughly 114.6 million people will be living with incurable blindness and 587.6 million people will be affected by severe visual impairment [8]. Although some affected individuals can be treated with surgery or medication, 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") may be the only option [18]. Analogous to cochlear implants, these devices electrically stimulate surviving cells in the visual pathway to evoke visual percepts ("phosphenes"). The phosphenes produced by current prostheses generally provide users with an improved ability to localize high-contrast objects and perform basic orientation & mobility tasks [1], but are not yet able to match the acuity of natural vision.

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 (roughly 500 retinal prostheses implanted to date). To investigate functional recovery and experiment with different implant designs, researchers have therefore been developing virtual reality (VR) prototypes that rely on simulated prosthetic vision (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 head-mounted display (HMD). This allows sighted participants to "see through the eyes" of the bionic eye user, taking into account their head and/or eye movements as they explore a virtual environment [27].

However, most SPV studies simply present stimuli on a computer monitor or an HMD, without taking into account eye movements, head motion, or locomotion [26]. This leads to a low level of immersion [25, 36], which refers to technical manipulations that separate

the existence of the physical world from the virtual world [34]. Seeing the real world, using a keyboard or joystick to move, and sounds present in the real environment are factors which lead to lower levels of immersion. However, the role of immersion for behavioral tasks in SPV is still unclear as no previous study has assessed whether behavioral performance is comparable across monitor-based and HMD-based versions of the same task.

In addition, most current prostheses provide a very limited field of view (FOV); for example, the artificial vision generated by Argus II [31], the most widely adopted retinal implant thus far, is restricted to roughly 10×20 degrees of visual angle. This forces users to scan the environment with strategic head movements while attempting to piece together the information [16]. The emergence of immersive VR as a research tool provides researchers with the ability to simulate this in a meaningful way.

Moreover, a recent literature review found that most SPV studies relied on phosphene models with a low level of biological realism [26]. It is therefore unclear how the findings of most SPV studies would translate to real bionic eye users.

To address these challenges, we make three contributions:

- We present VR-SPV, a virtual reality (VR) reality toolbox for simulated prosthetic vision (SPV) that allows sighted participants to "see through the eyes" of a bionic eye user.
- Importantly, we use an established and psychophysically validated computational model of bionic vision [4] to generate realistic SPV predictions.
- We systematically evaluate 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.

2 BACKGROUND

2.1 Restoring Vision with a Bionic Eye

The only FDA-approved technology for the treatment of retinal degenerative blindness are visual neuroprostheses. These devices consist of an electrode array implanted into the eye or brain that is used to artificially stimulate surviving cells in the visual system. Two retinal implants are already commercially available (Argus II: 60 electrodes, Second Sight Medical Products, Inc. [31]; Alpha-AMS (1600 electrodes, Retina Implant AG, [43]), and many other emerging devices have reached the clinical or pre-clinical stage [2, 17, 29]. In order to create meaningful progress within these devices, there is a growing need to understand how the vision these devices provide differs from natural sight [16].

One common fallacy is the assumption that each electrode produces a focal spot of light in the visual field [28, 39, 42]. 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 [6, 19, 32].

Increasing evidence suggests that perceived phosphene shape in an epiretinal implant is a result of unintended stimulation of nerve fiber bundles (NFBs) in the retina [6, 40]. These NFBs follow polar trajectories [23] away from the horizontal meridian, forming arch-like projections into the optic nerve (Fig. 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, *right*).

Ref. [6] demonstrated through simulations that the shape of elicited 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.

2.2 Related Work

The use of virtual reality has emerged as a tool for assisting users with low vision (see [26] for a review of recent literature). This includes not just accessibility aids, but also simulations aimed at understanding low vision. A number of previous SPV 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 [30, 49] or in HMDs [9, 52]. Some studies also tested the influence of FOV [41, 45] and eye gaze compensation [46] on acuity. Others focused on slightly more complex tasks such as letter [56], word [38], face [11, 14], and object recognition [33, 50, 55]. 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 of the studies mentioned above relied on the scoreboard model, it is unclear how their findings would
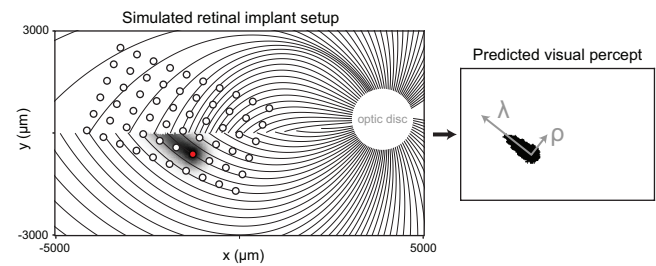
Figure 2: A simulated map of retinal NFBs (*left*) can account for visual percepts (*right*) elicited by retinal implants (reprinted with permission from [5]). *Left*: Electrical stimulation (red circle) of a NFB (black lines) could activate retinal ganglion cell bodies peripheral to the point of stimulation, 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.

translate to real bionic eye users. Although some studies attempted to address phosphene distortions [20, 44, 52], most did not account for the neuroanatomy (e.g., NFB trajectories) when deciding how to distort phosphenes. Only a handful of studies have incorporated a great amount of neurophysiological detail into their setup [24, 45, 49, 50], only two of which [45, 50] relied on an established and psychophysically validated model of SPV. One notable example is the study by Thorn *et al.* [45], which accounted for unintentional stimulation of axon fibers in the retina by adding a fixed "tail" length to each phosphene. However, a fixed-length tail is a simplification of the model [6] as the size of phosphenes (and their tails) have been shown to vary with stimulation parameters such as amplitude, frequency, and pulse duration [35].

In addition, being able to move around as one would in real life has shown to significantly increase the amount of immersion a user experiences [36]. However, the level of immersion offered by most SPV studies is relatively low, as stimuli are often presented on a screen [50, 53]. 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 [16]. 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, sometimes even exceeding theoretical acuity limits (as pointed out by [10]).

In summary, previous SPV research has assumed that the scoreboard model produces phosphenes that are perceptually similar to real bionic vision [48, 56], and that findings from an HMD-based task would more accurately represent the experience of a bionic eye user than a monitor version [41, 45, 54]. In this paper we aim to systematically evaluate these assumptions with a within-subjects (repeated measures) design, allowing for direct comparisons in performance across different model parameters and display conditions.

3 METHODS

3.1 VR-SPV: A Virtual Reality Toolbox for Simulated Prosthetic Vision

The VR-SPV system consisted of either a wireless head-mounted VR headset (HTC VIVE Pro Eye with wireless adapter, HTC Corporation) or a standard computer monitor (Asus VG248QE, 27in, 144Hz, 1920x1080p). Both HMD and monitor versions used the same computer for image processing (Intel i9-9900k processor and an Nvidia RTX 2070 Super GPU with 16GB of DDR4 memory). All software was developed using the Unity development platform, consisting of a combination of C# code processed by the central processing unit (CPU) and fragment/compute shaders processed by the graphics processing unit (GPU). The entire software package, along with a more in-depth explanation, is available at https://github.com/bionicvisionlab/BionicVisionXR.

The workflow for simulating bionic vision was as follows:

i. Image acquisition: Utilize Unity's virtual camera to acquire the scene at roughly 90 frames per second and downscale to a target texture of 86×86 pixels.

ii. Image processing: Conduct any image preprocessing specified by the user. Examples include grayscaling, extracting and enhancing edges, and contrast maximization. In the current study, the image was converted to grayscale and edges were extracted in the target texture with a 3×3 Sobel operator.

iii. Electrode activation: Determine electrode activation based on the visual input as well as the placement of the simulated retinal implant. In the current study, a 3×3 Gaussian blur was applied to the preprocessed image to average the grayscale values around each electrode's location in the visual field. This gray level was then interpreted as a current amplitude delivered to a particular electrode in the array.

iv. Phosphene model: Use Unity shaders to convert electrode activation information to a visual scene in real time. The current study re-implemented the axon map model available in *pulse2percept* [4] using shaders.

v. Phosphene rendering: Render the elicited phosphenes either on the computer monitor or the HMD of the VR system.

The VR-SPV system is designed to handle any retinal implant by allowing users to specify the location and size of each electrode in the simulated device. It can also handle other phosphene models, including cortical models, by replacing the model provided by *pulse2percept* with any phosphene model of their choosing.

While not considered in this study, VR-SPV can also be used to model temporal interactions by integrating electrode activation from previous frames or by only rendering at a specific frequency. The software is also capable of utilizing the VIVE's eye tracking hardware to elicit a "gaze lock". This function moves the rendered image to the center of the user's gaze, attempting to replicate the inability of a prosthetic user to scan the presented image with eye movements. Neither of these optional functions were used in this study as they were not the focus of the current work, and it was unclear how these settings would influence any findings on the parameters being studied. VR-SPV also includes a function to change the source of the visual input from a virtual environment to the HMD's front-facing camera for supporting AR applications.

3.2 Simulated Prosthetic Vision

The underlying phosphene model for this experiment was a re-implementation of the pyschophysically validated axon map model [6] provided by *pulse2percept* [4]. To support real-time execution, an initial mapping of each electrode's effects on the scene were pre-calculated with *pulse2percept* before starting the experiment. 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 2). As can be seen in Fig. 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 [6]. Although the reason for this is not fully understood, it is clear that the choice of these
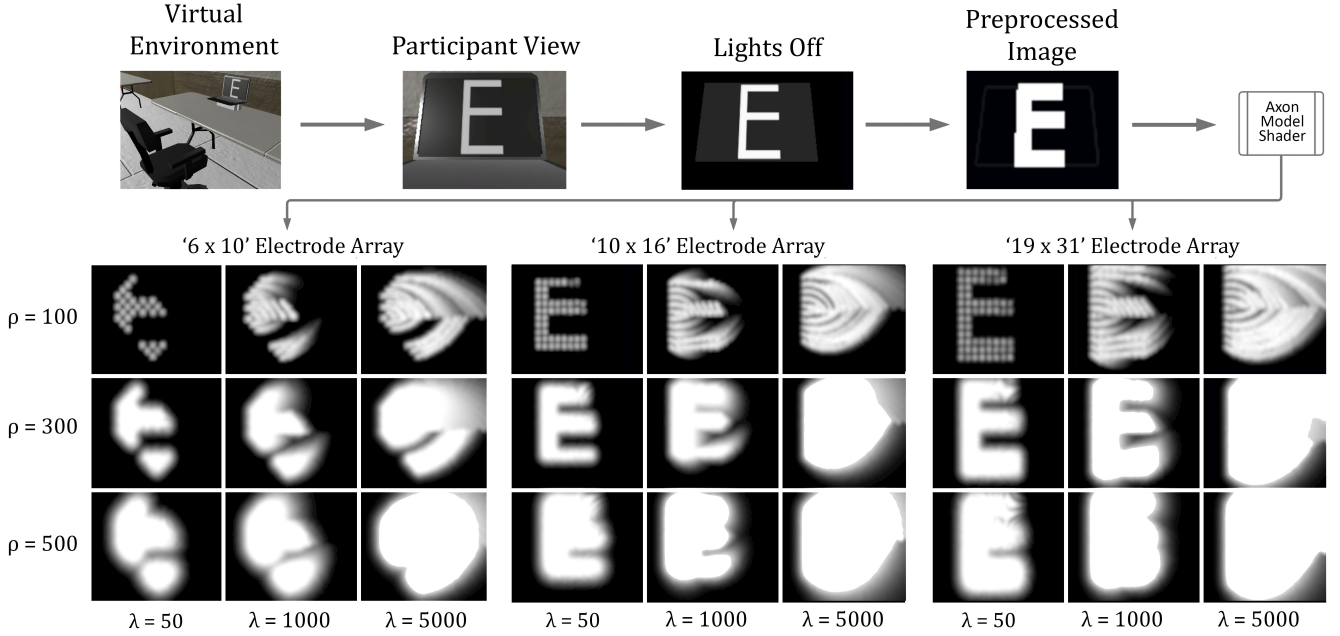
Figure 3: Letter recognition task. *Top:* **The lights in the virtual room are turned off and the image seen by the user is passed to the preprocessing shader which performs edge extraction/enhancement before the axon model shader renders SPV. Modeled after [13].** *Bottom:* **Output of the axon model shader across the various devices and ρ / λ combinations.**

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 [5].
- Argus IV (hypothetical): $19 \times 31 = 589$ electrodes spaced 575 µm apart in a rectangular grid implanted at 0°.

3.3 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.

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 µm) with the smallest possible phosphene size and the highest number of electrodes. The order of all subsequent blocks was randomized for each participant.

4 EXPERIMENTS AND RESULTS

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 [13], 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 [22].

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.

The procedures and results for each task are presented separately below, followed by a joint discussion on both experiments in the subsequent sections.
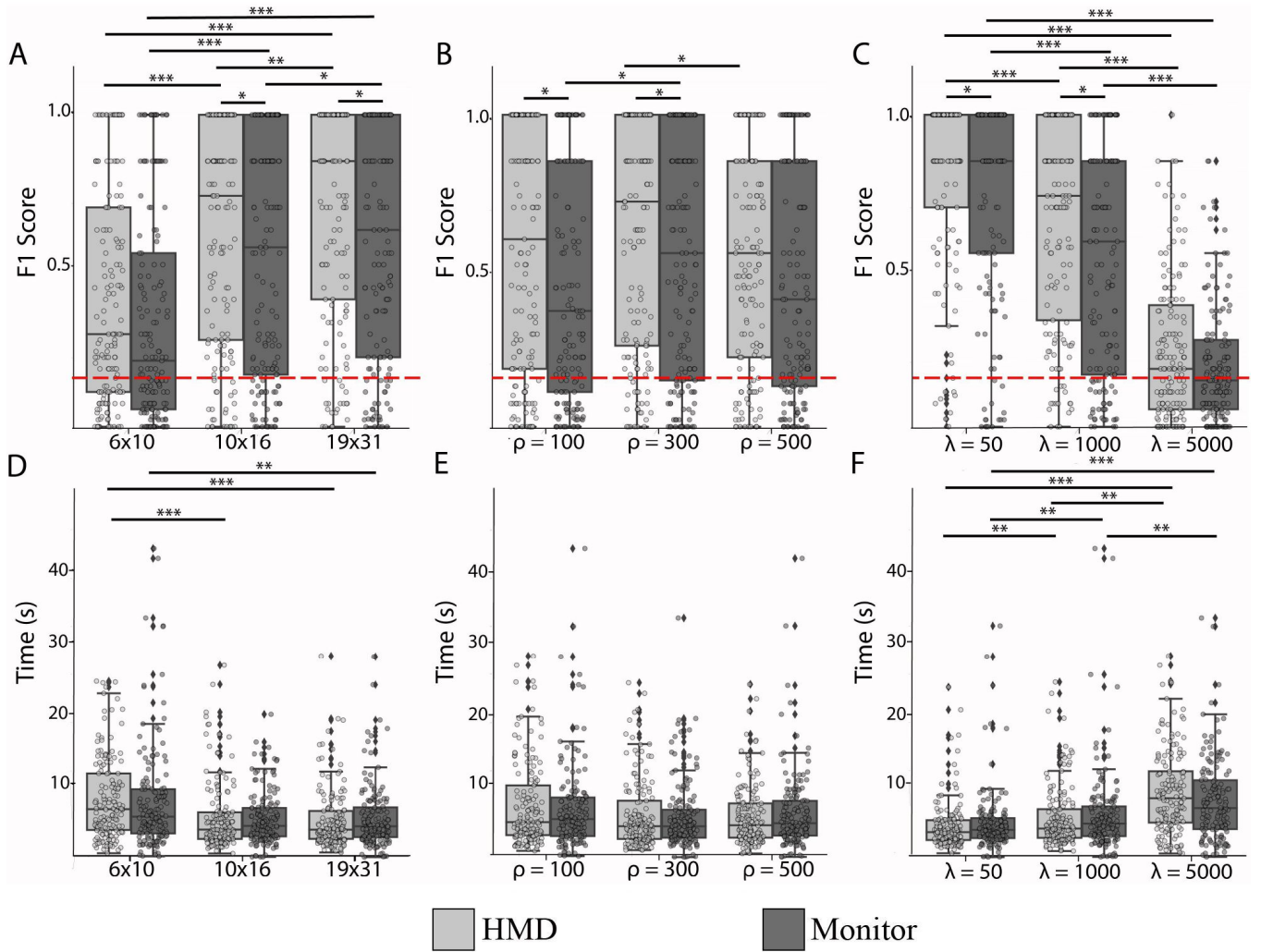
Figure 4: Letter recognition task. Data points represent each subject's average performance in a block with boxplots displaying median and interquartile ranges. *Top:* **Average F1 score across blocks for each subject within the condition specified by the x-axis.** *Bottom*: **Average time across blocks for each subject within the condition specified by the x-axis. Statistical significance was determined using ART ANOVA (***<.05, **<.01, ***<.001).**

4.1 Task 1: Letter Recognition

4.1.1 Original Task. The first experiment was modeled after a letter recognition task performed by Argus II recipients [13]. 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.

4.1.2 Experimental Setup and Procedure. To emulate the experiment described in [13], 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. 3).

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 × 3 ρ values × 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.

	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 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.

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 the VR-SPV toolbox was used to generate SPV. To mimic the training session of [13], 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.

4.1.3 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 [37]. 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 AR-Tool [51] 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 [51]. Posthoc analyses were performed on significant groups by analyzing the rank-transformed contrasts [15]. The Tukey method [47] 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.

4.1.4 Results. Results from the letter recognition task are summarized in Table 1 and distributions are plotted in Fig. 4. Group F-values, along with their significance, are reported in Table 2. Each data point in Fig. 4 represents a subject's F1 score (Fig. 4A–C) and

time per trial (Fig. 4D–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 ($\frac{1}{9} = 0.1111$).

As expected, increasing the number of electrodes (Fig. 3A) significantly increased F1 scores in both HMD (light gray) and monitor (dark gray) versions of the task. 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. 3D). However, increasing the number of electrodes from 10×16 to 19×31 did not further decrease recognition time.

Contrary to previous findings, F1 scores and recognition time did not systematically vary as a function of phosphene size (ρ, Fig. 3B, E). In both HMD and monitor-based conditions, median F1

	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 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.
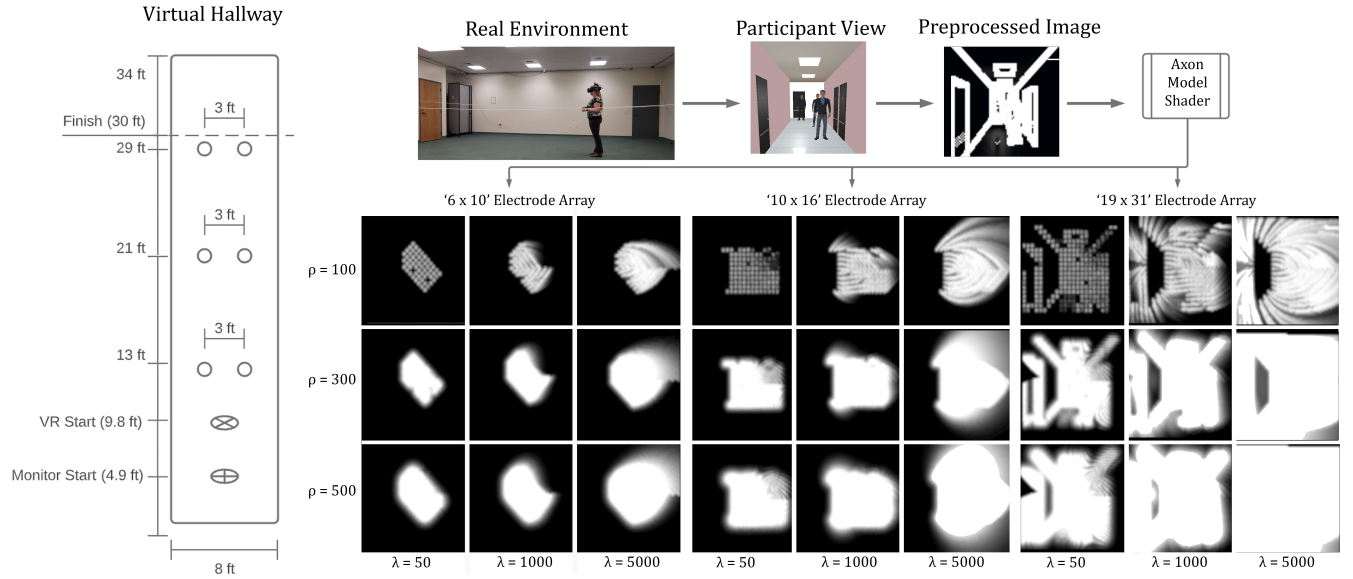
Figure 5: Obstacle avoidance task. *Left:* **Layout of the virtual hallway environment modeled after [22]. Empty circles represent the possible locations for obstacles.** *Right/Top:* **View of the real environment -> participant's view is passed to the preprocessing shader which performs edge extraction/enhancement before the axon model shader renders SPV.** *Bottom:* **Output of the axon model shader across the various devices and ρ / λ combinations.**

scores were highest for $\rho = 300\,\mu m$ (Table 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. 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 1). Increasing λ also significantly increased the time it took participants to identify the letter.

A trend toward a higher F1 score when using the HMD was observed across all conditions (Fig. 4, *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, *Top*). While average time per trial was faster across all conditions with the HMD, the effect was not significant (Fig. 3, *Bottom*).

4.2 Task 2: Obstacle Avoidance

4.2.1 Original Task. The second task was modeled after an obstacle avoidance experiment performed by Argus II recipients [22]. 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 trial, the experimenter instructed the participant to stop when they reached the end of the hallway.

4.2.2 Experimental Setup and Procedure. To emulate the experiment described in [22], we designed a virtual hallway (Fig. 5, *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. 5).

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 × 3 ρ values × 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 5, *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
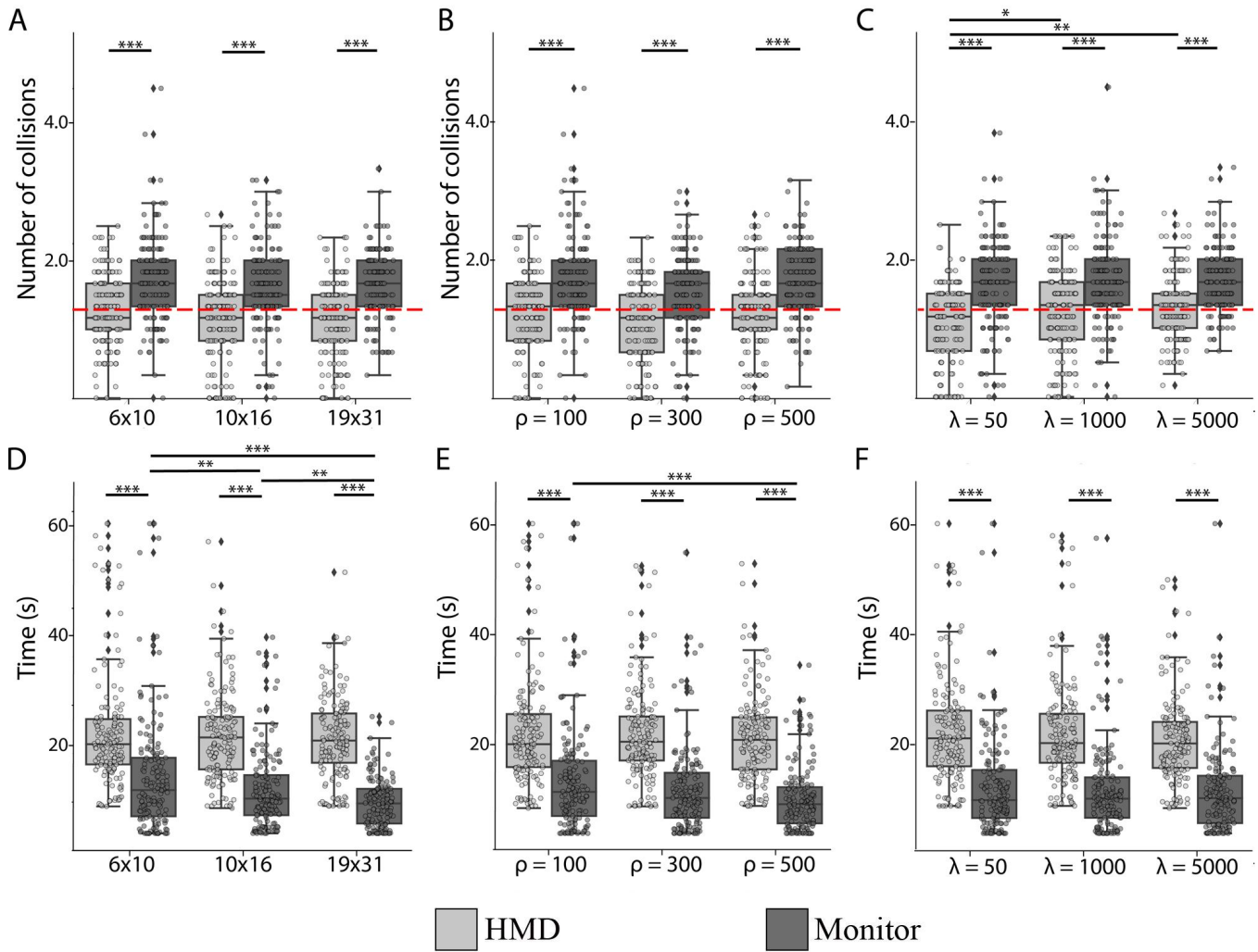
Figure 6: Obstacle avoidance. Data points represent each subject's average performance in a block with boxplots displaying median and interquartile ranges. *Top:* **Average number of collisions across blocks for each subject within the condition specified by the x-axis. Red line represents chance level (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).**

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 $0.7\,m \times 0.4\,m$ hitbox and the participant having a radius of $0.4\,m$. Subject locations and orientations were continuously recorded. Time per trial, along with individual positions and timings of each collision, were recorded for each trial.

4.2.3 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

| | 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 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.

significant groups using the Tukey method for multiple comparison adjustments.

4.2.4 Results. Results are summarized in Table 3 and Fig. 6. Each data point in Fig. 6 represents a subject's number of collisions (Fig. 6, *Top*) and time to completion (Fig. 6, *Bottom*) averaged across repetitions in a block. Group F-values, along with their significance, are reported in Table 4.

Contrary to our expectations, neither the number of electrodes (Fig. 6A) nor phosphene size (Fig. 6B) had a significant effect on the number of collisions. Although the number of collisions decreased slightly with higher electrode counts (Table 3), this did not reach statistical significance. The only statistical differences 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.

| | 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: 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.

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. 6D). Similarly to comparisons across groupings of ρ values, there was a slight downward trend across the median time taken as phosphene distortion increased (Fig. 6, *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. 6, *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. 6, *Bottom*).

5 DISCUSSION

5.1 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 1). However, this trend was only significant for the hypothetical future devices and smaller phosphene sizes and elongations (Fig. 4, *Top*). While average time per trial was faster across all conditions with the HMD, the effect was not significant (Fig. 3, *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. 6, *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.

5.2 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 1 and 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. 4C), which is much better than the behavioral metrics reported with real Argus II patients [13]. 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 [22].

Contrary to our expectations, phosphene size (ρ) did not systematically affect performance (Fig. 4B, Fig. 6B). 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 [12, 21]

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 [6], suggesting future work should seek to use psychophysically validated SPV models when making theoretical predictions about device performance.

5.3 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. 4A). This performance benefit was also observed in the time it took participants to recognize the letter (Fig. 4D), and is consistent with previous literature on object recognition [45].

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. 6A), 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 [3, 7].

5.4 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 [7].

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 [34], 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 and 6) which were not unlike those reported in the literature [13, 22]. 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.

6 CONCLUSIONS

The present work constitutes a first essential step towards immersive VR simulations of bionic vision. Data from two behavioral experiments demonstrate the importance of choosing an appropriate level of immersion and phosphene model complexity. The VR-SPV toolbox that enabled these experiments is freely available at https://github.com/bionicvisionlab/BionicVisionXR and designed to be extendable to a variety of bionic eye technologies. Overall this work has the potential to further our understanding of the qualitative experience associated with different bionic eye technologies and provide realistic expectations of prosthetic performance.

ACKNOWLEDGMENTS

REFERENCES

[1] Lauren N. Ayton, Nick Barnes, Gislin Dagnelie, Takashi Fujikado, Georges Goetz, Ralf Hornig, Bryan W. Jones, Mahiul M. K. Muqit, Daniel L. Rathbun, Katarina Stingl, James D. Weiland, and Matthew A. Petoe. 2020. An update on retinal prostheses. *Clinical Neurophysiology* 131, 6 (June 2020), 1383–1398. https://doi.org/10.1016/j.clinph.2019.11.029

[2] Lauren N. Ayton, Peter J. Blamey, Robyn H. Guymer, Chi D. Luu, David A. X. Nayagam, Nicholas C. Sinclair, Mohit N. Shivdasani, Jonathan Yeoh, Mark F. McCombe, Robert J. Briggs, Nicholas L. Opie, Joel Villalobos, Peter N. Dimitrov, Mary Varsamidis, Matthew A. Petoe, Chris D. McCarthy, Janine G. Walker, Nick Barnes, Anthony N. Burkitt, Chris E. Williams, Robert K. Shepherd, Penelope J. Allen, and for the Bionic Vision Australia Research Consortium. 2014. First-in-Human Trial of a Novel Suprachoroidal Retinal Prosthesis. *PLOS ONE* 9, 12 (Dec. 2014), e115239. https://doi.org/10.1371/journal.pone.0115239 Publisher: Public Library of Science.

[3] Matthew R. Behrend, Ashish K. Ahuja, Mark S. Humayun, Robert H. Chow, and James D. Weiland. 2011. Resolution of the Epiretinal Prosthesis is not Limited by Electrode Size. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19, 4 (Aug. 2011), 436–442. https://doi.org/10.1109/TNSRE.2011.2140132

[4] M. Beyeler, G. M. Boynton, I. Fine, and A. Rokem. 2017. pulse2percept: A Python-based simulation framework for bionic vision. In *Proceedings of the 16th Science in Python Conference*, K. Huff, D. Lippa, D. Niederhut, and M. Pacer (Eds.). 81–88. https://doi.org/10.25080/shinma-7f4c6e7-00c

[5] Michael Beyeler, Geoffrey M. Boynton, Ione Fine, and Ariel Rokem. 2019. Model-Based Recommendations for Optimal Surgical Placement of Epiretinal Implants. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019 (Lecture Notes in Computer Science)*, Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan (Eds.). Springer International Publishing, 394–402.

[6] Michael Beyeler, Devyani Nanduri, James D. Weiland, Ariel Rokem, Geoffrey M. Boynton, and Ione Fine. 2019. A model of ganglion axon pathways accounts for percepts elicited by retinal implants. *Scientific Reports* 9, 1 (June 2019), 1–16. https://doi.org/10.1038/s41598-019-45416-4

[7] Michael Beyeler, Ariel Rokem, Geoffrey M. Boynton, and Ione Fine. 2017. Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies. *Journal of Neural Engineering* 14, 5 (Aug. 2017), 051003. https://doi.org/10.1088/1741-2552/aa795e Publisher: IOP Publishing.

[8] Rupert R. A. Bourne, Seth R. Flaxman, Tasanee Braithwaite, Maria V. Cicinelli, Aditi Das, Jost B. Jonas, Jill Keeffe, John H. Kempen, Janet Leasher, Hans Limburg, Kovin Naidoo, Konrad Pesudovs, Serge Resnikoff, Alex Silvester, Gretchen A. Stevens, Nina Tahhan, Tien Y. Wong, Hugh R. Taylor, Rupert Bourne, Peter Ackland, Aries Arditi, Yaniv Barkana, Banu Bozkurt, Tasanee Braithwaite, Alain Bron, Donald Budenz, Feng Cai, Robert Casson, Usha Chakravarthy, Jaewan Choi, Maria Vittoria Cicinelli, Nathan Congdon, Reza Dana, Rakhi Dandona, Lalit Dandona, Aditi Das, Iva Dekaris, Monte Del Monte, Jenny Deva, Laura Dreer, Leon Ellwein, Marcela Frazier, Kevin Frick, David Friedman, Joao Furtado, Hua Gao, Gus Gazzard, Ronnie George, Stephen Gichuhi, Victor Gonzalez, Billy Hammond, Mary Elizabeth Hartnett, Minguang He, James Hejtmancik, Flavio Hirai, John Huang, April Ingram, Jonathan Javitt, Jost Jonas, Charlotte Joslin, Jill Keeffe, John Kempen, Moncef Khairallah, Rohit Khanna, Judy Kim, George Lambrou, Van Charles Lansingh, Paolo Lanzetta, Janet Leasher, Jennifer Lim,

Hans Limburg, Kaweh Mansouri, Anu Mathew, Alan Morse, Beatriz Munoz, David Musch, Kovin Naidoo, Vinay Nangia, Maria Palaiou, Maurizio Battaglia Parodi, Fernando Yaacov Pena, Konrad Pesudovs, Tunde Peto, Harry Quigley, Murugesan Raju, Pradeep Ramulu, Serge Resnikoff, Alan Robin, Luca Rossetti, Jinan Saaddine, Mya Sandar, Janet Serle, Tueng Shen, Rajesh Shetty, Pamela Sieving, Juan Carlos Silva, Alex Silvester, Rita S. Sitorus, Dwight Stambolian, Gretchen Stevens, Hugh Taylor, James Tielsch, Miltiadis Tsilimbaris, Jan van Meurs, Rohit Varma, Gianni Virgili, Jimmy Volmink, Ya Xing Wang, Ning-Li Wang, Sheila West, Peter Wiedemann, Tien Wong, Richard Wormald, and Yingfeng Zheng. 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 (Sept. 2017), e888–e897. https://doi.org/10.1016/S2214-109X(17)30293-0 Publisher: Elsevier.

[9] Xiaofei Cao, Heng Li, Zhuofan Lu, Xinyu Chai, and Jing Wang. 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)*. 1–5. https://doi.org/10.1109/CISP-BMEI.2017.8302244

[10] Avi Caspi and Ari Z. Zivotofsky. 2015. Assessing the utility of visual acuity measures in visual prostheses. *Vision Research* 108 (March 2015), 77–84. https://doi.org/10.1016/j.visres.2015.01.006

[11] M. H. Chang, H. S. Kim, J. H. Shin, and K. S. Park. 2012. Facial identification in very low-resolution images simulating prosthetic vision. *Journal of Neural Engineering* 9, 4 (July 2012), 046012. https://doi.org/10.1088/1741-2560/9/4/046012 Publisher: IOP Publishing.

[12] S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell. 2009. Simulating prosthetic vision: I. Visual models of phosphenes. *Vision Research* 49, 12 (June 2009), 1493–506.

[13] Lyndon da Cruz, Brian F. Coley, Jessy Dorn, Francesco Merlini, Eugene Filley, Punita Christopher, Fred K. Chen, Varalakshmi Wuyyuru, Jose Sahel, Paulo Stanga, Mark Humayun, Robert J. Greenberg, Gislin Dagnelie, and for the Argus II Study Group. 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 (May 2013), 632–636. https://doi.org/10.1136/bjophthalmol-2012-301525 Publisher: BMJ Publishing Group Ltd Section: Clinical science.

[14] Grégoire Denis, Christophe Jouffrais, Victor Vergnieux, and Marc Macé. 2013. Human faces detection and localization with simulated prosthetic vision. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 61–66. https://doi.org/10.1145/2468356.2468368

[15] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 754–768. https://doi.org/10.1145/3472749.3474784

[16] Cordelia Erickson-Davis and Helma Korzybska. 2021. What do blind people "see" with retinal prostheses? Observations and qualitative reports of epiretinal implant users. *PLOS ONE* 16, 2 (Feb. 2021), e0229189. https://doi.org/10.1371/journal.pone.0229189 Publisher: Public Library of Science.

[17] Laura Ferlauto, Marta Jole Ildelfonsa Airaghi Leccardi, Naïg Aurelia Ludmilla Chenais, Samuel Charles Antoine Gilliéron, Paola Vagni, Michele Bevilacqua, Thomas J. Wolfensberger, Kevin Sivula, and Diego Ghezzi. 2018. Design and validation of a foldable and photovoltaic wide-field epiretinal prosthesis. *Nature Communications* 9, 1 (March 2018), 1–15. https://doi.org/10.1038/s41467-018-03386-7

[18] Eduardo Fernandez. 2018. Development of visual Neuroprostheses: trends and challenges. *Bioelectronic Medicine* 4, 1 (Aug. 2018), 12. https://doi.org/10.1186/s42234-018-0013-8

[19] I. Fine and G. M. Boynton. 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 (Sept. 2015), 20140208. https://doi.org/10.1098/rstb.2014.0208

[20] Xiaoli Guo, Zheng Jin, Xinyang Feng, and Shanbao Tong. 2014. Enhanced Effective Connectivity in Mild Occipital Stroke Patients With Hemianopia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22, 6 (Nov. 2014), 1210–1217. https://doi.org/10.1109/TNSRE.2014.2325601 Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[21] Nicole Han, Sudhanshu Srivastava, Aiwen Xu, Devi Klein, and Michael Beyeler. 2021. Deep Learning–Based Scene Simplification for Bionic Vision. In *Augmented Humans Conference 2021*. ACM, Rovaniemi Finland, 45–54. https://doi.org/10.1145/3458709.3458982

[22] Yingchen He, Susan Y. Sun, Arup Roy, Avi Caspi, and Sandra R. Montezuma. 2020. Improved mobility performance with an artificial vision therapy system using a thermal sensor. *Journal of Neural Engineering* 17, 4 (Aug. 2020), 045011. https://doi.org/10.1088/1741-2552/aba4fb Publisher: IOP Publishing.

[23] Nomdo M. Jansonius, Julia Schiefer, Jukka Nevalainen, Jens Paetzold, and Ulrich Schiefer. 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 (Dec. 2012), 70–78. https://doi.org/10.1016/j.exer.2012.10.008

[24] Horace Josh, Collette Mann, Lindsay Kleeman, and Wen Lik Dennis Lui. 2013. Psychophysics testing of bionic vision image processing algorithms using an FPGA Hatpack. In *2013 IEEE International Conference on Image Processing*. 1550–1554. https://doi.org/10.1109/ICIP.2013.6738319 ISSN: 2381-8549.

[25] Suzan (Suzie) Kardong-Edgren, Sharon L. Farra, Guillaume Alinier, and H. Michael Young. 2019. A Call to Unify Definitions of Virtual Reality. *Clinical Simulation in Nursing* 31 (June 2019), 28–34. https://doi.org/10.1016/j.ecns.2019.02.006

[26] Justin Kasowski, Byron A. Johnson, Ryan Neydavood, Anvitha Akkaraju, and Michael Beyeler. 2021. Furthering Visual Accessibility with Extended Reality (XR): A Systematic Review. *arXiv:2109.04995 [cs]* (Sept. 2021). http://arxiv.org/abs/2109.04995 arXiv: 2109.04995.

[27] Justin Kasowski, Nathan Wu, and Michael Beyeler. 2021. Towards Immersive Virtual Reality Simulations of Bionic Vision. In *Augmented Humans Conference 2021*. ACM, Rovaniemi Finland, 313–315. https://doi.org/10.1145/3458709.3459003

[28] Jessica Kvansakul, Lachlan Hamilton, Lauren N. Ayton, Chris McCarthy, and Matthew A. Petoe. 2020. Sensory augmentation to aid training with retinal prostheses. *Journal of Neural Engineering* 17, 4 (July 2020), 045001. https://doi.org/10.1088/1741-2552/ab9e1d

[29] H. Lorach, G. Goetz, R. Smith, X. Lei, Y. Mandel, T. Kamins, K. Mathieson, P. Huie, J. Harris, A. Sher, and D. Palanker. 2015. Photovoltaic restoration of sight with high visual acuity. *Nat Med* 21, 5 (May 2015), 476–82. https://doi.org/10.1038/nm.3851

[30] Yanyu Lu, Panpan Chen, Ying Zhao, Jingru Shi, Qiushi Ren, and Xinyu Chai. 2012. Estimation of Simulated Phosphene Size Based on Tactile Perception. *Artificial Organs* 36, 1 (2012), 115–120. https://doi.org/10.1111/j.1525-1594.2011.01288.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1525-1594.2011.01288.x.

[31] Yvonne Hsu-Lin Luo and Lyndon da Cruz. 2016. The Argus® II Retinal Prosthesis System. *Progress in Retinal and Eye Research* 50 (Jan. 2016), 89–107. https://doi.org/10.1016/j.preteyeres.2015.09.003

[32] Yvonne H-L. Luo, Joe Jiangjian Zhong, Monica Clemo, and Lyndon da Cruz. 2016. Long-term Repeatability and Reproducibility of Phosphene Characteristics in Chronically Implanted Argus II Retinal Prosthesis Subjects. *American Journal of Ophthalmology* 170 (Oct. 2016), 100–109. https://doi.org/10.1016/j.ajo.2016.07.021

[33] Marc J.-M. Macé, Valérian Guivarch, Grégoire Denis, and Christophe Jouffrais. 2015. Simulated Prosthetic Vision: The Benefits of Computer-Based Object Recognition and Localization. *Artificial Organs* 39, 7 (2015), E102–E113. https://doi.org/10.1111/aor.12476

[34] Haylie L. Miller and Nicoleta L. Bugnariu. 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 (April 2016), 246–256. https://doi.org/10.1089/cyber.2014.0682 Publisher: Mary Ann Liebert, Inc., publishers.

[35] Devyani Nanduri, Ione Fine, Alan Horsager, Geoffrey M. Boynton, Mark S. Humayun, Robert J. Greenberg, and James D. Weiland. 2012. Frequency and Amplitude Modulation Have Different Effects on the Percepts Elicited by Retinal Stimulation. *Investigative Ophthalmology & Visual Science* 53, 1 (Jan. 2012), 205–214. https://doi.org/10.1167/iovs.11-8401

[36] Marco Pasch, Nadia Bianchi-Berthouze, Betsy van Dijk, and Anton Nijholt. 2009. Immersion in Movement-Based Interaction. In *Intelligent Technologies for Interactive Entertainment*, Anton Nijholt, Dennis Reidsma, and Hendri Hondorp (Eds.). Vol. 9. Springer Berlin Heidelberg, Berlin, Heidelberg, 169–180. https://doi.org/10.1007/978-3-642-02315-6_16 Series Title: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering.

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

[38] Angelica Perez Fornos, Jörg Sommerhalder, and Marco Pelizzone. 2011. Reading with a Simulated 60-Channel Implant. *Frontiers in Neuroscience* 5 (2011), 57. https://doi.org/10.3389/fnins.2011.00057

[39] Alejandro Perez-Yus, Jesus Bermudez-Cameo, Jose J. Guerrero, and Gonzalo Lopez-Nicolas. 2017. Depth and Motion Cues with Phosphene Patterns for Prosthetic Vision. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. 1516–1525. https://doi.org/10.1109/ICCVW.2017.179 ISSN: 2473-9944.

[40] Joseph F. Rizzo, John Wyatt, John Loewenstein, Shawn Kelly, and Doug Shire. 2003. Perceptual Efficacy of Electrical Stimulation of Human Retina with a Microelectrode Array during Short-Term Surgical Trials. *Investigative Ophthalmology & Visual Science* 44, 12 (Dec. 2003), 5362–5369. https://doi.org/10.1167/iovs.02-0817 Publisher: The Association for Research in Vision and Ophthalmology.

[41] Melani Sanchez-Garcia, Ruben Martinez-Cantin, Jesus Bermudez-Cameo, and Jose J. Guerrero. 2020. Influence of field of view in visual prostheses design: Analysis with a VR system. *Journal of Neural Engineering* 17, 5 (Oct. 2020), 056002. https://doi.org/10.1088/1741-2552/abb9be Publisher: IOP Publishing.

[42] Melani Sanchez-Garcia, Ruben Martinez-Cantin, and Josechu J. Guerrero. 2019. Indoor Scenes Understanding for Visual Prosthesis with Fully Convolutional Networks. In *VISIGRAPP*. https://doi.org/10.5220/0007257602180225

[43] K. Stingl, K. U. Bartz-Schmidt, D. Besch, A. Braun, A. Bruckmann, F. Gekeler, U. Greppmaier, S. Hipp, G. Hortdorfer, C. Kernstock, A. Koitschev, A. Kusnyerik, H. Sachs, A. Schatz, K. T. Stingl, T. Peters, B. Wilhelm, and E. Zrenner. 2013. Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS. *Proc Biol Sci* 280, 1757 (April 2013), 20130077. https://doi.org/10.1098/rspb.2013.0077

[44] Mahadevan Subramaniam, Parvathi Chundi, Abhilash Muthuraj, Eyal Margalit, and Sylvie Sim. 2012. Simulating prosthetic vision with disortions for retinal prosthesis design. In *Proceedings of the 2012 international workshop on Smart health and wellbeing (SHB '12)*. Association for Computing Machinery, New York, NY, USA, 57–64. https://doi.org/10.1145/2389707.2389719

[45] Jacob Thomas Thorn, Enrico Migliorini, and Diego Ghezzi. 2020. Virtual reality simulation of epiretinal stimulation highlights the relevance of the visual angle in prosthetic vision. *Journal of Neural Engineering* 17, 5 (Nov. 2020), 056019. https://doi.org/10.1088/1741-2552/abb5bc Publisher: IOP Publishing.

[46] Samuel A. Titchener, Mohit N. Shivdasani, James B. Fallon, and Matthew A. Petoe. 2018. Gaze Compensation as a Technique for Improving Hand–Eye Coordination in Prosthetic Vision. *Translational Vision Science & Technology* 7, 1 (Jan. 2018), 2. https://doi.org/10.1167/tvst.7.1.2

[47] John W. Tukey. 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics* 5, 2 (1949), 99–114. https://doi.org/10.2307/3001913 Publisher: [Wiley, International Biometric Society].

[48] Victor Vergnieux, Marc J.-M. Macé, and Christophe Jouffrais. 2014. Wayfinding with simulated prosthetic vision: Performance comparison with regular and structure-enhanced renderings. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2585–2588. https://doi.org/10.1109/EMBC.2014.6944151 ISSN: 1558-4615.

[49] Milena Vurro, Anne Marie Crowell, and John S. Pezaris. 2014. Simulation of thalamic prosthetic vision: reading accuracy, speed, and acuity in sighted humans. *Frontiers in Human Neuroscience* 8 (2014), 816. https://doi.org/10.3389/fnhum.2014.00816

[50] Lihui Wang, Fariba Sharifian, Jonathan Napp, Carola Nath, and Stefan Pollmann. 2018. Cross-task perceptual learning of object recognition in simulated retinal implant perception. *Journal of Vision* 18, 13 (Dec. 2018), 22. https://doi.org/10.1167/18.13.22

[51] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 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*. ACM, Vancouver BC Canada, 143–146. https://doi.org/10.1145/1978942.1978963

[52] Hao Wu, Jing Wang, Heng Li, and Xinyu Chai. 2014. Prosthetic vision simulating system and its application based on retinal prosthesis. In *2014 International Conference on Information Science, Electronics and Electrical Engineering*, Vol. 1. 425–429. https://doi.org/10.1109/InfoSEEE.2014.6948145

[53] Zhao Ying, Geng Xiulin, Li Qi, and Jiang Guangqi. 2018. Recognition of virtual maze scene under simulated prosthetic vision. In *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*. 1–5. https://doi.org/10.1109/ICACI.2018.8377543

[54] Marc P. Zapf, Paul B. Matteucci, Nigel H. Lovell, Steven Zheng, and Gregg J. Suaning. 2014. Towards photorealistic and immersive virtual-reality environments for simulated prosthetic vision: Integrating recent breakthroughs in consumer hardware and software. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2597–2600. https://doi.org/10.1109/EMBC.2014.6944154 ISSN: 1558-4615.

[55] Ying Zhao, Yanyu Lu, Yukun Tian, Liming Li, Qiushi Ren, and Xinyu Chai. 2010. Image processing based recognition of images with a limited number of pixels using simulated prosthetic vision. *Information Sciences* 180, 16 (Aug. 2010), 2915–2924. https://doi.org/10.1016/j.ins.2010.04.021

[56] Ying Zhao, Yanyu Lu, Ji Zhao, Kaihu Wang, Qiushi Ren, Kaijie Wu, and Xinyu Chai. 2011. Reading Pixelized Paragraphs of Chinese Characters Using Simulated Prosthetic Vision. *Investigative Ophthalmology & Visual Science* 52, 8 (July 2011), 5987–5994. https://doi.org/10.1167/iovs.10-5293