Influence of field of view in visual prostheses design: Analysis with a VR system

Melani Sanchez-Garcia*, Ruben Martinez-Cantin, Jesus Bermudez-Cameo, Jose J. Guerrero

Instituto de Investigación en Ingeniería de Aragón, (I3A). Universidad de Zaragoza, Spain

E-mail: `mesangar@unizar.es`

Abstract.

Visual prostheses are designed to restore partial functional vision in patients with total vision loss. Retinal visual prostheses provide limited capabilities as a result of low resolution, limited field of view and poor dynamic range. Understanding the influence of these parameters in the perception results can guide prostheses research and design. In this work, we evaluate the influence of field of view with respect to spatial resolution in visual prostheses, measuring the accuracy and response time in a search and recognition task. Twenty-four normal participants were asked to find and recognize usual objects, such as furniture and home appliance in indoor room scenes. For the experiment, we use a new simulated prosthetic vision system that allows simple and effective experimentation. Our system uses a virtual-reality environment based on panoramic scenes. The simulator employs a head-mounted display which allows users to feel immersed in the scene by perceiving the entire scene all around. Our experiments use public image datasets and a commercial head-mounted display. We have also released the virtual-reality software for replicating and extending the experimentation. Results show that the accuracy and response time decrease when the field of view is increased. Furthermore, performance appears to be correlated with the angular resolution, but showing a diminishing return even with a resolution of less than 2.3 phosphenes per degree. For the design of retinal protheses, it is better to concentrate the phosphenes in a small area, to maximize the angular resolution, even if that implies sacrificing field of view.

Keywords: Simulated prosthetic vision, retinal prostheses, low vision, computer vision, field of view, virtual-reality system, object recognition.

1. Introduction

Retinitis pigmentosa and age-related macular degeneration are two important causes of visual impairment and blindness [1, 2]. These diseases involve gradual loss of photoreceptor, or rod cells, while generally preserving the inner retinal cells. This results in a progressive loss of vision. Retinal prostheses are a promising technology
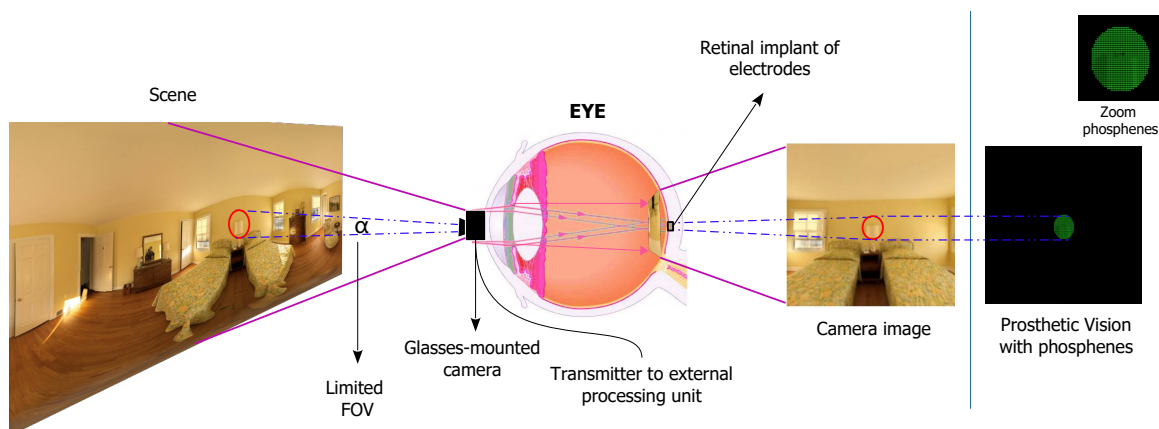
Figure 1. Overview of a retinal prosthesis. The external and internal components include a glasses-mounted camera, an external processing unit and an implanted electrode array. First, the external camera acquires an image. Then, the external processor converts the image to a suitable pattern of electrical stimulation of the retina through an electrode array. The result is a phosphene image with limited field of view (FOV).

to improve vision in patients with such advanced degenerative diseases [3–6]. These visual prostheses can partially restore vision, bypassing damaged photoreceptors and electrically stimulating the surviving retinal cells, such as the retinal ganglion cells [7]. The Argus II epiretinal prostheses (Second Sight Medical Products Inc., Sylmar, CA, USA) is the most widely used retinal prostheses world-wide [3,6]. It is made up of an external component (glasses-mounted camera) and an implanted component (electrode implant), as can be seen in Figure 1. The camera acquires images from the real-world that are transmitted to a portable visual processing unit linked to the camera. The processed information is sent to the retina via electrical impulses in the implant by an electrode array. The stimulation can activate a group of neurons in a small localized area of the retina leading to percepts of spots of light known as "phosphenes" [8]. The brain interprets patterns of phosphenes in the restricted area as visual information. Results in implanted subjects have demonstrated partial visual restoration, with improvement in both coarse objective function and performance of everyday tasks [9–18].

There are still physiological and technological limitations of the information received by implanted patients. The number of electrodes and implant size limit the maximum amount of information that can be provided by the stimulating array. This fact has restricted the degree of visual resolution (up to 1500 phosphenes) and dynamic range of the visual perception (8 gray levels) that can be delivered to the user. Depth perception is also not possible due to low resolution or monocular implantation. Besides, current systems provides a field of view (FOV) of approximately $18° \times 11°$ in the retinal area, which correspond to the FOV subtended by the electrode implant on the retina, as can be seen in Figure 1.

The restricted FOV impairs functional vision regarding to mobility or recognition

reducing quality of life [19–21]. To improve the design of current visual prostheses in terms of FOV and resolution, one of the questions to be answered is whether expanding the FOV of the input image can benefit implanted subjects. Several devices for retinitis pigmentosa patients (tunnel vision) have been designed to expand the projected FOV on their retina, for example, minimizing the scene zooming out using a camera [22, 23]. He et al. [24] studied the influence of FOV in the Argus II system with thermal imaging by changing the mapping (zoom out) between the sensor and the electrode array. Zooming out turns out in a larger visual angle mapped onto the same implant region on the retina, effectively decreasing the spatial resolution of their prosthetic vision. Alternatively, Ameri et al. [25] have desgined an implant with wider FOV, by spreading the electrodes in a wider retinal surface. This setup seem to have advantages for motion perception and head scanning, at the expense of a reduced electrode density and fine detail. Therefore, the optimal implant area for a fixed number of electrodes remains an open question, which we address in this study.

One of the concerns that has limited the development and wider use of prosthetic devices is how to evaluate their utility and function in terms of benefit to the patient and, consequently, how to predict in which direction to develop these devices [26]. Most studies in this field have incorporated performance-based measures and questionnaires to try to understand the relative importance of visual parameters such as resolution, FOV and visual acuity, in the performance of daily tasks in subjects with visual impairment. However, studies with implanted subjects have limited statistical power and require cumbersome experiments. For example, the study of He et al. [24] was limited to four implanted participants. Simulated prosthetic vision (SPV) opens the opportunity to evaluate potential and forthcoming functionality, in early stages of design, of these implants with larger studies by using normal participants. SPV is a standard procedure for non-invasive evaluation using normal vision subjects. Furthermore, SPV systems allow for quickly or interactive modifications of the parameters of the simulation.

Researchers have previously used SPV for analysis of the visual perception in terms of resolution or FOV with normal sight subjects. For example, Fornos et al. [27] used SPV to study how such restrictions of the amount of visual information provided would affect performance on simple pointing and manipulation tasks. Hayes et al. [28] designed a set of tasks to assess performance of object recognition and manipulation and reading using different sizes of electrode array. Contrary to our approach in this paper, they used a constant electrode density, meaning that wider arrays also had larger number of electrodes. In a posterior study, Dagnelie et al. [29] explored minimal visual resolution requirements of a simulated retinal electrode array for mobility in real and virtual environments, experienced by normally sighted subjects in video headsets.

These experiments relied on some form of head-mounted device which allowed a virtual-reality experience during the experiment, but the *immersiveness* of experience was poor due to the technological limitations compared to modern commercial VR systems. Furthermore, most SPV studies use computer screens to present phosphenic images [30–32]. Modern commercial VR systems have specifications in terms of
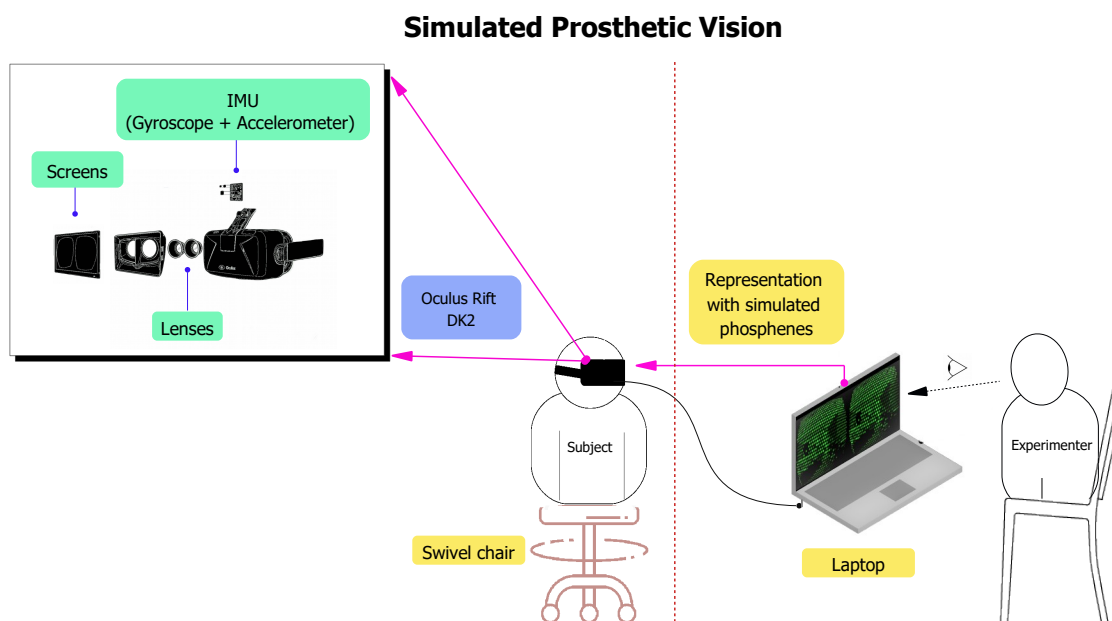
Simulated Prosthetic Vision

Figure 2. SPV system. The components consist of an Oculus Rift powered by a consumer level laptop. The VR system is composed by two lenses, two screens and a suite of internal sensors (gyroscope and accelerometer). The representation with simulated phosphenes is displayed on the laptop screen as well as on the Oculus system worn by the subjects. During the experiment, subjects were seated in a swivel chair allowing them to scan the entire scene all around them (360 degrees).

resolution, latency and response time that allows a fully immersive experience with unexpensive equipment and computers. Using these devices, subjects are immersed and able to interact with complex environments. SPV with commercial VR systems, such as the Oculus Rift used in this work, can be a useful tool to evaluate everyday tasks in more realistic setups and complex simulations. Furthermore, prosthetic vision can be assessed in controlled, real or simulated environments.

In this work, we evaluated the influence of FOV and resolution of the prostheses on the subject's performance in a recognition task, since it is of high priority for patients with visual diseases such as retinitis pigmentosa. Concretely, we analyzed an object search and recognition task performance in indoor scenes with different reduced FOVs and resolutions limited to hundreds of electrodes. For that, we present a new VR system for more realistic SPV environments using panoramic scenes. The VR system could be extended to more complex tasks because it supports realistic environments. This system acts as an electronic visual aid that attach to the user's head and presents information directly to the user's eyes. The panoramic scenes allow subjects felt immersed in the scene by perceiving the entire scene (360 degrees).
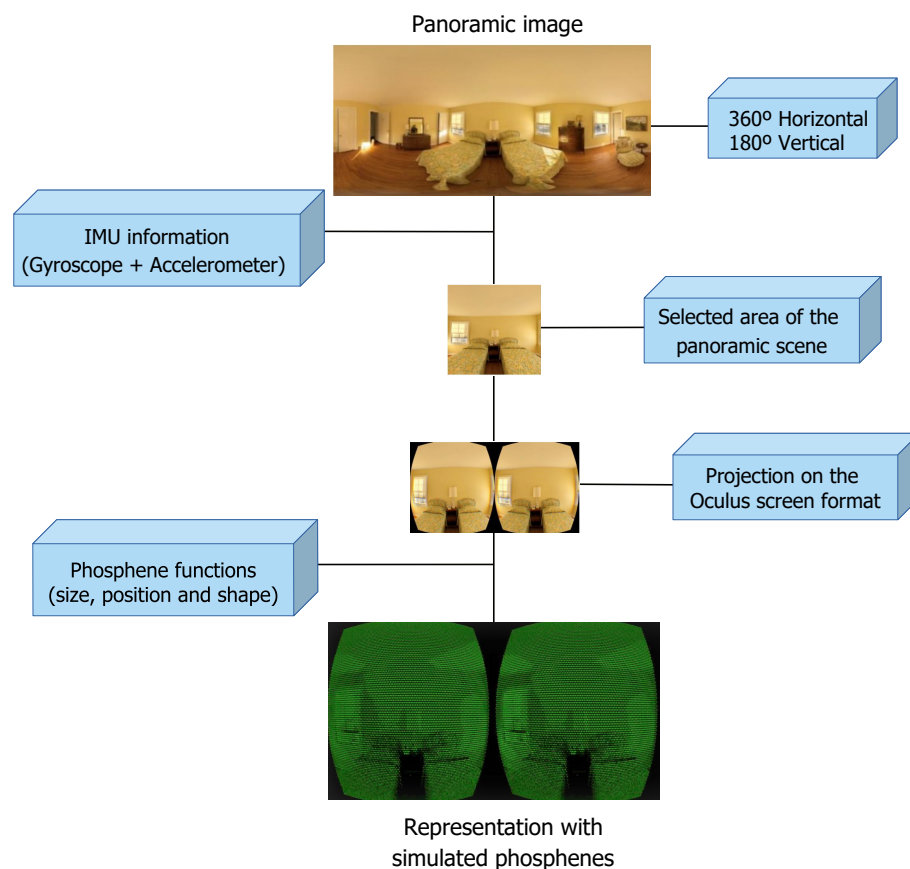
Figure 3. Data process. The input scene in our VR system is a panoramic image in equirectangular representation. The system estimate the head orientation using the IMU sensors (gyroscope and accelerometer) allowing to choose the area of the panoramic scene that is being observed at the moment. The area selected is then projected on the two Oculus screens and represented using simulated phosphenes.

2. Materials and Methods

We evaluate the influence of FOV on object search and recognition tasks using SPV through a VR system. The SPV system is a standard procedure for non-invasive evaluation using normal vision subjects. This methodology allows controlled evaluation of normally sighted subject response and task performance which is fundamental to know the way humans perceive and interpret phosphenized renderings. SPV also offers the advantage of adapting implant designs to improve the perceptual quality without involving implanted subjects.

2.1. Participants

Twenty four subjects aged 20-30 (6 female, 18 male) participated in the experiment. They had no visual problems or wore their normal optical correction during the experiment.
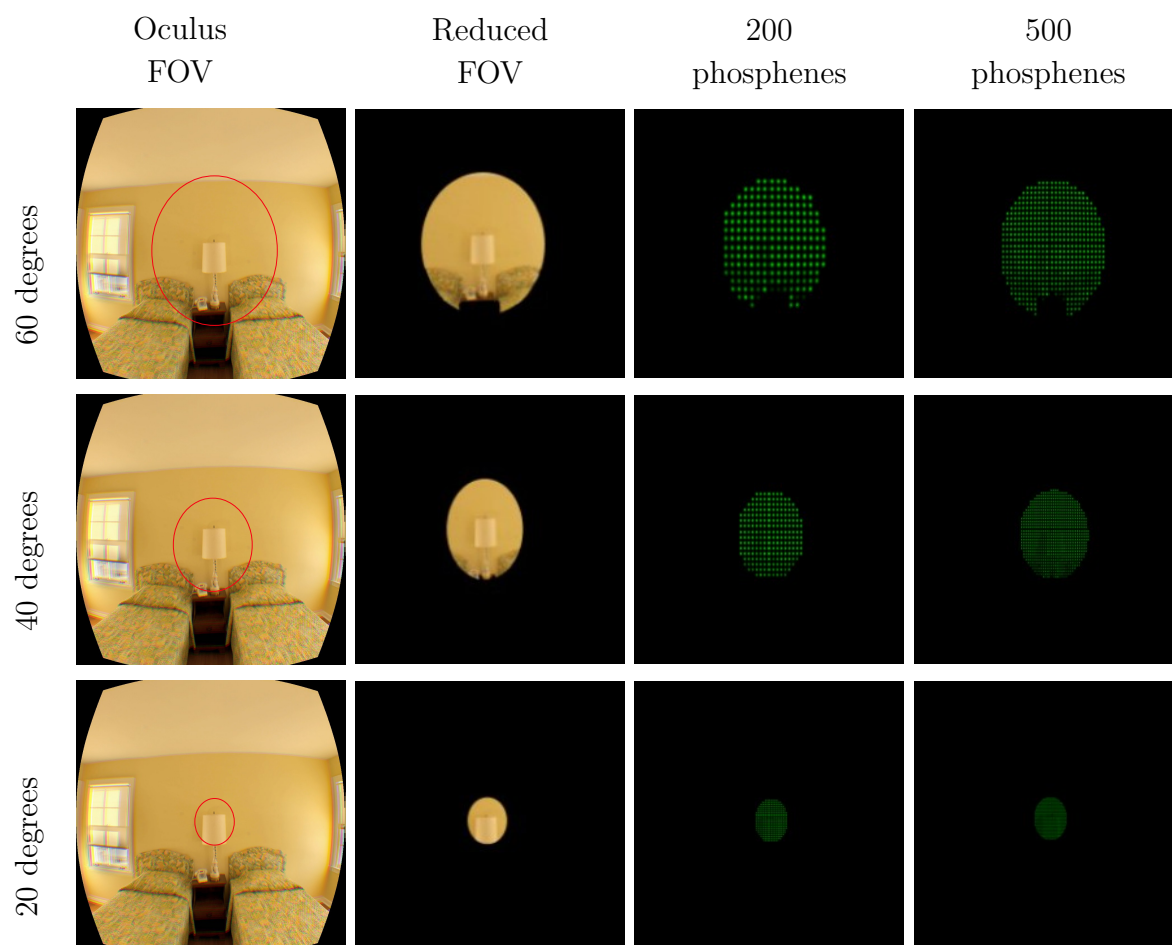
Figure 4. Stimuli conditions in the experiment. Rows: different FOVs used in the experiment (60, 40 and 20 degrees). Columns: different resolutions used in the experiment (200 and 500 phosphenes). Note that in the last row only part of the lamp is visible, the beds cannot be seen without moving the point of view.

2.2. Simulated prosthetic vision (SPV)

This section describes the SPV system including the hardware specifications, software components and phosphene generation.

2.2.1. Hardware. The experiment was conducted on an Oculus Rift powered by a consumer level laptop (Intel(R) Core(TM) i5-8265U CPU), but this system is capable of working in real time with any mid-range laptop (see Figure 2). The VR system, Oculus Rift DK2, is composed by two lenses, two screens and an inertial measurement unit (IMU) with gyroscopes and accelerometers, a standard setup for most commercial VR systems. It contains 5.7 inch dual OLED screens with a resolution of 960 x 1080 projected on to each eye. Each display is projected into the eye using a lens with pincushion distortion to provide peripheral vision. In our experiments we mostly use the central part of the display which remains undistorted. The representation with simulated phosphenes was displayed on the VR system worn by the participants as

Figure 5. Object classes considered during the experiment. Subjects have to recognize the main objects of the scene, those objects that are usually present in hotel rooms such as bed, window, chair, tv, sofa, table/nightstand, door, wardrobe/shelving and lamp.

well as on the laptop screen for the experimenter to check the progress. During the experiment, participants were seated in a swivel chair allowing them to scan the entire scene with head rotation movements.

2.2.2. Software. The implementation was done in C++, using the Oculus Rift SDK to connect with the VR system and OpenCV for image processing. Our software is compatible with the Windows operating system. Figure 3 shows the data process designed to generate the stimuli for the VR system. Starting from a panoramic scene of 360° of horizontal FOV and 180° of vertical FOV, the system estimates the orientation of the user using the information collected from the IMU (Figure 2). The selected area is then projected on the two Oculus displays and finally represented with simulated phosphenes. The projection models for the VR system can be found in the *Appendix A*.

Our phosphene map configuration is similar to the frameworks of McKone et al. [33] and Chen et al. [34]. We approximate the phosphenes as green circular dots with a Gaussian luminance profile –each phosphene has maximum intensity at the center and gradually decays to the periphery, following a Gaussian function–. The intensity of a phosphene is directly extracted from the intensity of the same region in the image. For our experiments, each phoshene has 8 intensity levels. The size and brightness are directly proportional to the quantified sampled pixel intensities. The phosphene map is calculated and updated with respect to head orientation in real time. The complete process of phosphene generation can be found in the *Appendix B*.

The software used in the experiment to simulate the prosthetic vision with the VR system is available at `http://webdiis.unizar.es/~rmcantin/index.php/Research/Vrfov`.

2.3. Procedure

The experiment was conducted using a selection of indoor panoramic scenes from a public database of Xiao et al. [35] that are adapted to our SPV system. The resolution of the panoramic scenes is 1024 x 512 pixels. All the scenes from the database are

hotel rooms containing objects such as beds, tables, chairs, windows and doors, among others (see Figure 5). We removed several scenes because they had external distractions, such as signatures and watermarks, that could affect the experiment. From them, we randomly selected 50 scenes. The scenes were presented to the subjects using different stimuli conditions based on two resolutions (200 and 500 phosphenes) and three FOVs (20, 40 and 60 degrees), as can be seen in Figure 4. We selected these particular resolutions and FOVs based on current visual prostheses [26], although our VR platform allows to quickly change those parameters. The total number of stimuli scenes generated for the experiment was 300.

For the formal experiment, participants were recruited to complete an object search and recognition task using the SPV system. By turning the swivel chair and head, subjects had to scan the entire scene by changing the head orientation, but not the position. At the same time, participants had to search and recognize the main objects of the scene such as *bed, window, chair, tv, sofa, table/nighstand, door, wardrobe/shelving and lamp* (see Figure 5). Small objects such as telephones, vases or remote controls were not taken into account. We took into account both the type of recognized object and the number of each of them. At the beginning of the experiment, subjects were trained during 2 minutes on a test set of images. Participants were informed that all scenes were indoor scenes, but they were not informed about the class and number of objects in each scene nor the different stimuli conditions. The demo images were not included in the experiment to avoid learning effects.

Figure 6 shows the trial setup. Each trial consisted of a sequence of fifteen scenes generated by choosing a resolution, a FOV and a scene, among the shuffled conditions. Initially, a white dot was displayed in the center of the screen indicating where the participants had to maintain fixation until the beginning of the task. Next, each scene was presented and the user has a maximum of 60 seconds to scan the scene moving the chair and their head. The following scene was displayed when the participant verbally ordered that the task had been completed ("I cannot see more objects") or after 60 seconds. This procedure was repeated for each scenario of the trial. Between two scenes the participants were instructed to "look for" the white dot to fixate the gaze in a steady position before the next scene. There was a break time in the middle of the sequence (between $7^{th} - 8^{th}$ scene) of 60 seconds.

The participants verbally indicated the type of objects seen in each scene during the scanning to minimize distractions. The responses of each stimulus scene were timed recorded and annotated by the experimenter. If the participants did not respond within the 60 seconds frame, the result of that scene was considered as "no object was recognized" and a time of 60 seconds was recorded. Participants did not get feedback of their responses. The complete experiment took approximately 20 minutes per participant.
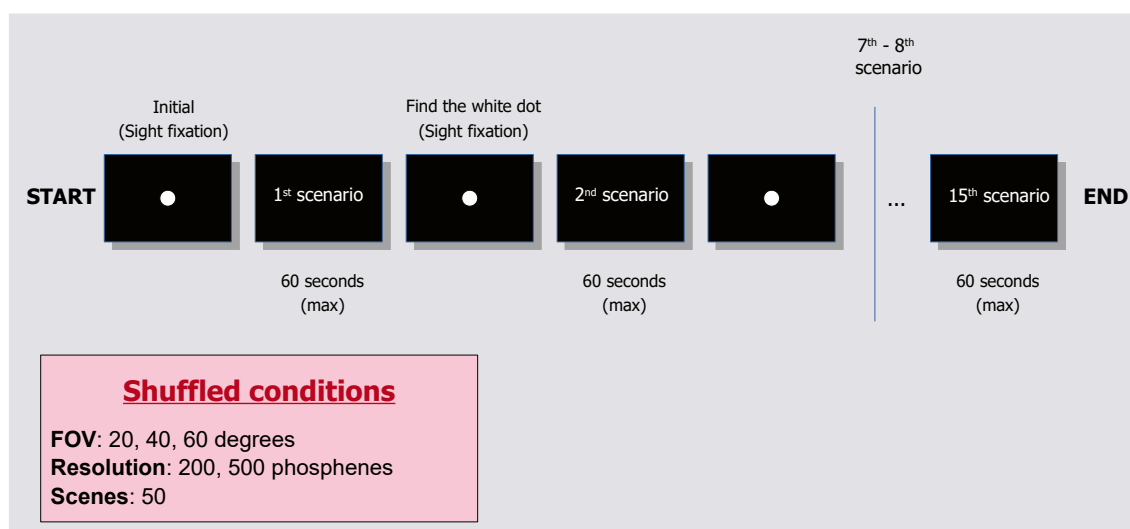
Figure 6. Trial setup. To generate the scene in each step of the trial sequence we used shuffled conditions of FOV, resolution and scenes. During the experiment, each scene appeared for 60 seconds and switched for the next scene automatically. Break time in the middle of the sequence (between $7^{th} - 8^{th}$ scene) was 60 seconds. The complete experiment took approximately 20 minutes.

2.4. Statistical analysis

Data were analyzed using two-way ANOVA and post hoc-test with Tukey's method to evaluate simultaneously the effect of the two grouping variables (resolution and FOV) on the response variables object recognition and recognition time with $p = 0.05$, * < 0.05 , ** < 0.01, *** < 0.001 and *ns* not significant.

3. Results

The performance for all resolutions and FOVs is summarized in Figure 7. The results show the ratio of recognized object and recognition time (mean ± standard deviation) for aggregated data from all subjects and all images. For the same scene, we normalized the results of the object recognition with all the conditions of the experiment, being 1 the experiment case with the most number recognized objects for a particular scene. The time recorded in each scene was the time it took the subject to recognize the first object. We also performed a test to determine if the mean difference between specific pairs of conditions are statistically significant using Tukey's method with a significant level $\alpha = 0.05$.

Figure 7(a) shows the object recognition performance for the three FOVs (20, 40 and 60 degrees) and the two resolutions (200 and 500 phosphenes). For the same resolution, the object recognition performance decreases as the FOV increases, which might seem initially counterintuitive. For the resolution of 500 phosphenes the average performance is 80.66±7.85, 63.84±8.68 and 34.74±7.83 for 20, 40 and 60 degrees respectively. For
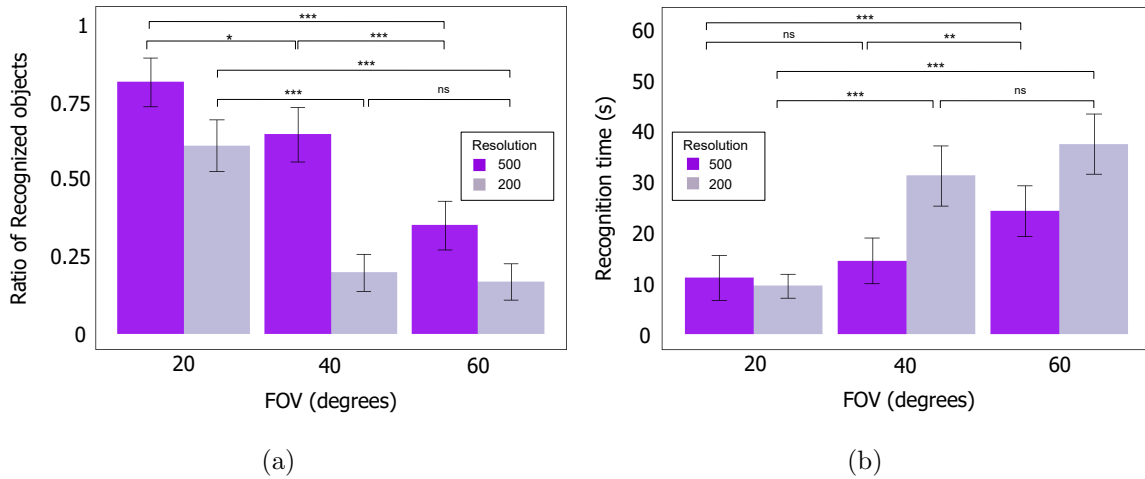
(a) (b)

Figure 7. Ratio of recognized objects and recognition time. (a) Object recognition results for 20, 40 and 60 FOVs and for the two resolutions, 200 and 500 phosphenes. High scores indicate that subjects were able to recognize most of the objects in each scene. (b) Recognition time results for 20, 40 and 60 FOVs and for the two resolutions, 200 and 500 phosphenes. High scores indicate that the subjects needed more time to perform the recognition task. ***=p<.001; **=p<.01; *=p<.05; ns=p>.05.

the resolution of 200 phosphenes the average performance is 60.24 ± 8.28, 19.65 ± 5.97, 16.58 ± 5.82 for 20, 40 and 60 degrees respectively. No significant differences were found for 40-60 FOVs for 200 phosphenes (p=0.8056), as performance was very poor in both cases. There was no significant difference between the two resolutions (p>0.05). Comparing the performance for the same FOV, the performance increases with the number of phosphenes for all cases.

Figure 7(b) shows the recognition time for the three FOVs and two resolutions. For the same resolution, the recognition time increases as the FOV increases. For the resolution of 500 phosphenes the average recognition time is 10.81 ± 4.32, 14.10 ± 4.41 and 23.71 ± 4.91 for 20, 40 and 60 degrees respectively. For the resolution of 200 phosphenes the average recognition time is 9.24 ± 2.35, 30.50 ± 5.79, 36.61 ± 5.77 for 20, 40 and 60 degrees respectively. No significant differences were found for 20-40 FOVs for 500 phosphenes (p=0.733) and similarly to 40-60 FOVs for 200 resolution (p=0.199). There was no significant difference between the two resolutions (p>0.05). Comparing the performance for the same FOV, the recognition time decreases with the number of phosphenes. Table 1 shows the number of cases considered failure, where no object has been recognized in 60s (>60s). The condition 60 FOV and 200 resolution had more number of failures. Contrary, the condition 20 FOV and 200 resolution had fewer number of failures.

Figure 8 shows box-plots of data distribution with 25, 50 and $75th$ quartiles for recognized objects and recognition time. In the recognized object (see Figure 8(a)), there is difference between the two resolution (200 and 500 phosphenes) for 20 and 40 FOVs. On the other hand, for 200 phosphenes there is only difference between 20 and
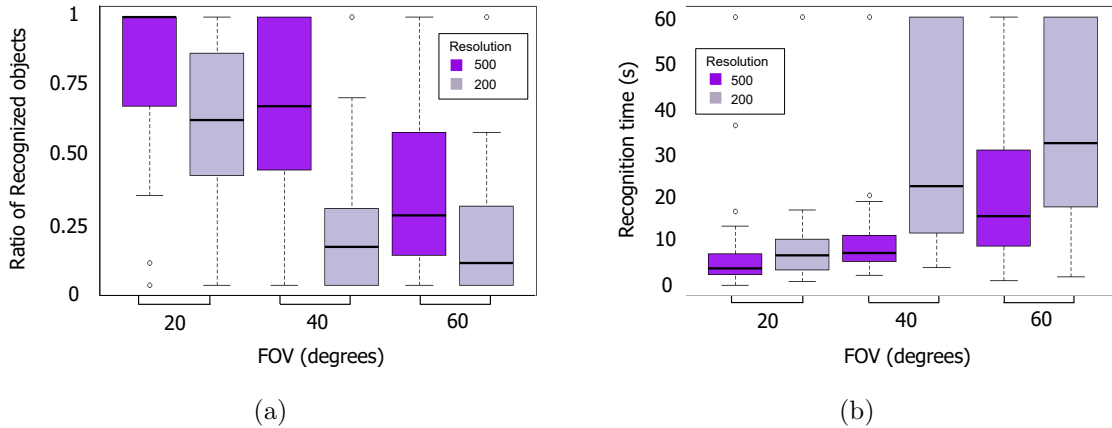
Figure 8. Ratio of recognized objects and recognition time. a) Box-plot for object recognition and b) box-plot for recognition time.

40 FOVs. In the case of recognition time (see Figure 8(b)), there is difference between the two resolution for 40 and 60 FOVs. At the same time, there is difference between 40 and 60 FOVs for 500 phosphenes and between 20 and 40 FOVs for 200 phosphenes. Note that the data distribution for the condition 20 FOV and 200 resolution and the condition 40 FOV and 500 resolution are very similar, even though they are two totally different schemes. However, it makes sense because they have practically the same angular resolution (see Figure 9).

Figure 9 shows the ratio of recognized objects and time recognition according to *angular resolution* (AR), measured in phosphenes per radian. We perform a logarithmic regression for all data. The logarithmic regression equation for the *object recognition rate* (OR) is $OR = -1.0345 + 0.4482 \cdot \log(AR)$ with $R^2 = 0.4031$ and $F - value = 201.3$. The positive coefficient indicates that as the angular resolution increases, the object recognition ratio tends to increase (see Figure 9(a)). The logarithmic regression equation for the *recognition time* (RT) is $RT = 83.14 - 18.78 \cdot \log(AR)$ with $R^2 = 0.2344$ and $F - value = 91.26$. The negative coefficient indicates that as the angular resolution increases, the recognition time tends to decrease (see Figure 9(b)). The regression output shows that both, ratio of recognized object and the recognition time variables are statistically significant with $p - value < 0.05$ (see Table 1).

4. Discussion

Recent developments in the technology for retinal implants should allow to improve the quality of electrode arrays, in terms of features, such as response time, resolution or size [25, 36]. Thus, an open question for the prosthesis design is how to evaluate and predict the utility and functionality in terms of patient benefit with respect to design parameters [26]. Several studies have incorporated performance-based measures and
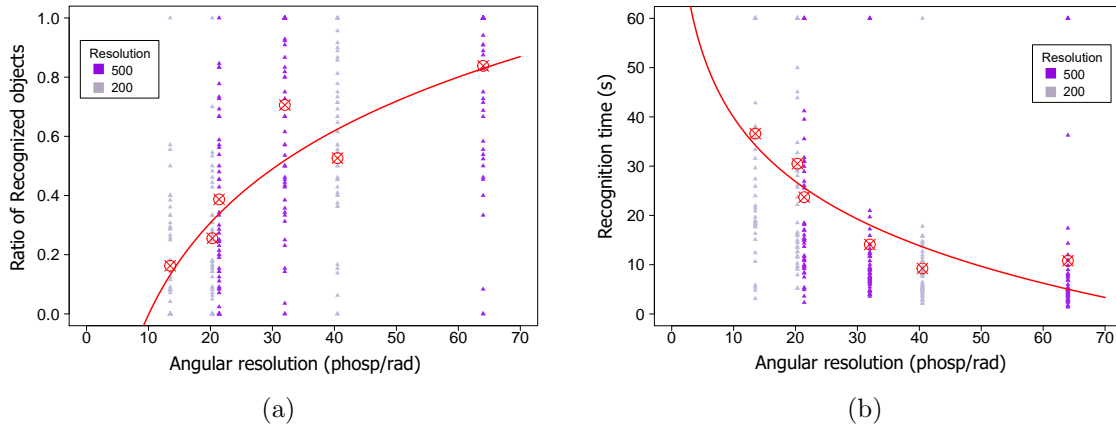
(a)   (b)

Figure 9. Angular resolution. (a) Performance of recognized objects according to angular resolution (phosphenes/radian) and (b) recognition time according to angular resolution. We perform a logarithmic regression for all data. The performance of recognized object tends to increase with angular resolution. Contrary, the recognition time tends to decrease with the angular resolution.

Table 1. Mean value of the ratio of recognized object and recognition time for the six angular resolutions. Parameters for logarithmic regression and R^2, $F-value$ and $p-value$.

Angular resolution (phosp/rad)	Ratio of Recognized object	Recognition time (s)	# Results > 60s
13.5	0.166 ± 0.058	36.61 ± 5.77	20
20.3	0.197 ± 0.059	30.50 ± 5.79	14
21.4	0.346 ± 0.078	23.71 ± 4.91	7
32.0	0.638 ± 0.087	14.10 ± 4.41	4
40.5	0.602 ± 0.083	9.24 ± 2.35	1
64.0	0.807 ± 0.076	10.81 ± 4.32	4
intercept	−1.0345	83.14	−
slope	0.4482	−18.78	−
R^2	0.4031	0.2344	−
$F-value$	201.3	91.26	−
$p-value$	$* * *$	$* * *$	−

questionnaires to understand the relation of visual parameters in the performance of everyday tasks in visually impaired subjects.

In this work, we study how certain features interact and affect perception. Intuitively, wider sensors should improve perception by allowing a larger field of view which can help detect motion and perform scanning [25]. Our results seem to indicate that it is not the case for head scanning, if it comes at the expense of phosphene density. Similarly, He et al. [24] investigated the effect of a wider FOV of a retinal prostheses on the performance of Argus II users in an object localization task using a thermal sensor.

The results showed that users were able to find objects using the current $11° \times 18°$ FOV with no zoom with higher precision and speed than when using a wider FOV by zooming out the sensor input. They also suggest that a higher spatial resolution may be preferred over a wider input FOV. This has demonstrated the importance of some parameters such as FOV, visual resolution and angular resolution [37]. Others works have also shown that the lack of visual information caused by the low resolution and the restriction in a large portion of the FOV can be compensated by scene scanning [38, 39]. During scene scanning, the relevant information of the environment is actively sought, quickly and efficiently. For instance, there is some evidence on humans that eye and head rotation facilitates the learning process to recognize simple objects [38, 40–42].

Simulated prosthetic vision (SPV) can be used to estimate the visual requirements to deliver a sufficient visual resolution and FOV with a large statistical power and cost efficiency. It provides an opportunity for simulation-based research regarding the design of a functional visual prosthesis, improvement of functional vision in low-phosphene-count devices, and also as a tool to search for image processing strategies to impart the most understandable prosthetic vision making experimentation with implanted patients unnecessary. In the work of He at al. [24], the FOV is modified at the external sensor by introducing a zoom lens, therefore changing the perceived scale. In contrast, thanks to the use of SPV we can actually change the properties of the electrode array, altering the FOV at the stimulus, mainting the 1:1 scale ration between perception and real world. Visual scanning has also been studied in SPV showing that it has a positive impact on the task performance by allowing subjects to increase the visual information [43]. Cai et al. [44] noticed during experiments that head rotation also allowed subjects to expand their effective FOVs and rudimentary depth perception through parallax. Chen et al. [45] also reported that maximizing visual information with scanning techniques improves visual acuity performance.

The theoretical resolution achievable by present-day retinal implants such as Argus II is $4°$ [46]. There have also been cases reported in Argus II clinical trials with $1.1°$, well below the theoretical limit. This may be explained by effective scanning techniques, allowing subjects to temporally integrate percepts. Previous studies in SPV found out that a minimum angular resolution of 170 pixels/radian is needed for an acceptable accuracy in some tasks such as mobility and object recognition [27–29, 47]. However, this angular resolution is around three times more than the maximum angular resolution obtained in our experiment (64.0 phosphenes/radian).

We found that object recognition is well achieved with low resolution and restricted FOV. As can be seen in Figure 7(a), we obtained a significant improvement in the recognition performance as the FOV was reduced, for both resolutions. Besides, participants took less time to recognize objects with a narrower FOV (see Figure 7(b)). This seems counterintuitive since with a narrower FOV the global reference of the scene is lost. However, the narrower the FOV, the higher the angular resolution and therefore the greater the image detail (higher frequencies). Contrary, the widest FOV allows to cover the widest area of the scene but it only allows to see the gist of the image (lower

frequencies).

Psychophysical and computational studies have shown that high and low spatial frequency provide different content from a scene: higher spatial frequencies contain fine information of image details and/or object boundaries, whereas lower frequencies preserve coarse blobs representing the gist of the scene [48–51]. In our experiments, the narrower FOV produces a higher phosphene density or angular resolution. By increasing the FOV (20, 40, 60 degrees), the angular resolution decreases: 40.5, 20.3 and 13.5 phosphenes/radian for 200 phosphenes and 64.0, 32.0 and 21.4 phosphenes/radian for 500 phosphenes, as can be seen in Figure 9. For lower angular resolutions, there were more cases of responses from subjects that exceeded the 60 s limit in the experiment, which implies that the subjects need more time to recognize details in these conditions (see Table 1). Our results suggest that it is better to see higher spatial frequencies of the scene, even if the global reference is lost due to the narrow FOV, since the subjects are able to holding back the global concept of the scene through visual scanning. Further, in more complex scenes such as low-contrast or low-luminance level, the time required to recognize the objects was higher and recognition performance decreased. In the same way, in those scenes with high contrast the subjects needed less time and the performance in the task increased. This fact has been demonstrated by Ehrlich et al. [42], where they observed that as the contrast increases the perception of the scene increases.

Virtual-reality (VR) systems have been widely used for experimentation [52–56]. They improve the quality of SPV tests since the experimentation environment is closer to the real-world. Some SPV research such as Denis et al. [52] used a VR system with two videos cameras mounted on the front of the headset to capture the visual scene for text localization. In [53], the SPV was generated on a VR headset with a stereo camera that captured the scene in real-time for object recognition and localization. In similar studies, participants interacted with the VR system for different tasks such as visual acuity measurement [54] or wayfinding task [55]. Vergnieux [56] used a virtual environment that was displayed via a VR system for navigation task evaluation.

Despite these proposals, we need more realistic and easier to use experimentation environments to improve the quality of SPV tests. A complete immersive experience implies to consider rotations and translations in the virtual environment. This approach would require the development of a detailed 3D model of the scene, to track the user and a large space allowing the free movement of the user in the real world. Since the influence of visual scanning in the FOV is dominated by rotations we have considered a simpler option. The environment is represented by a 360 panoramic scene so that the head can be turned in all directions from the center of the scene and the subject can explore the entire scene. This approach has a set of advantages. Panoramic scenes are easy to obtain in any real scene. By contrast, modelling a complete 3D model of the environment is very expensive. Further, the rotational movement of the head can be easily obtained from the embedded inertial sensors of the headset avoiding external cameras for the translation tracking. The experiment can be performed sitting on a chair, avoiding collisions, the requirement of a very large space and facilitating repeatability. Furthermore, our SPV

system allows replication of the experiment using many subjects with normal vision without being limited by the number of implanted subjects.

5. Conclusions

Effective visual prosthesis design presents a challenge for engineers and neuroscientists. We have analyzed the influence of field of view with respect to resolution in visual prostheses through a study with a novel simulated prosthetic vision setup: a virtual-reality system using panoramic scenes. Participants perceived phosphene images in an immersive head-mounted device. The task consisted on finding and identifying common objects with different field of view and number of phosphenes. Our results show that, for the same number of phosphenes, recognition accuracy and response time improved by reducing the field of view. In fact, angular resolution is major determinant for effective object recognition, being directly correlated to the accuracy and inversely correlated to the response time. However, it has also shown a diminishing return even for an angular resolution of less than 2.3 phosphenes per degree. Simulated prosthetic vision allows experiments with larger number of participants and simpler procedures than clinical studies with implanted patients. Our experimental setup relies on a consumer-level head-mounted display, public image databases and we have released the software needed to run the simulator, to facilitate replications and extensions. Our results seem to indicate that it is better to concentrate the phosphenes to maximize the angular resolution, sacrificing field of view.

Ethical statement

The research process was conducted according to the ethical recommendations of the Declaration of Helsinki. The research protocol used for this study is non-invasive, purely observational, with absolutely no-risk for any participant. There was no personal data collection or treatment and all subjects were volunteers. Subjects gave their informed written consent after explanation of the purpose of the study and possible consequences. The consent allowed the abandonment of the study at any time. All data were analyzed anonymously. The experiment was approved by the Aragon Autonomous Community Research Ethics Committee (CEICA).

Appendix A

Spherical image projection in the VR system

Since the participants only rotate the head and body during the experiment, in our SPV system, we only consider rotation movements. In the absence of translation, all the visual information of the environment can be encoded in a spherical image, i.e. an equirectangular panorama. Thus, if the rotation of the head is known, the image received by the subject through the screen can be virtually simulated from a panoramic image.

To achieve the effect of the stimulus image on the VR system we combine two different projections models: the spherical projection that models the projection on the panoramic image and the perspective projection that models the projection on the viewfinder screen.

In the Oculus VR system, the inertial measurement unit (IMU) collect the information of the head movement made by the user in form of quaternions like $q = a + bi + cj + dk$. This information is used to generate the rotation matrix which describes the basis of the absolute reference system used by the IMU with respect to the screen reference system as:

$$R_{imu} = \begin{pmatrix} 1 - b^2 - c^2 & 2ab - 2cd & 2ac + 2bd \\ 2ab + 2cd & 1 - 2a^2 - 2c^2 & 2cb + 2ad \\ 2ac - 2bd & 2bc + 2ad & 1 - 2a^2 - 2b^2 \end{pmatrix} \tag{A.1}$$

This rotation matrix defines the head orientation and, therefore, it can be used to find the area of the panoramic scene that is currently being observed. Then, the selected area is projected on the screen.

For the generation of the screen image, we calculate the ray corresponding to each phosphene on the screen as

$$\vec{v} = R_{pan}^T R_{imu}^T K^{-1} u \tag{A.2}$$

where $\vec{v} = (v_x, v_y, v_z)^T$ is the direction vector of the ray in the panorama reference system, R_{imu} is the rotation matrix from the IMU data, R_{pan} is the the basis of the panorama reference system with respect to the absolute reference system of the IMU and $u = (j, i, 1)^T$ are the coordinates of each phosphene on the screen image. K is the calibration matrix used in the perspective projection that models the projection in the screen.

The spherical coordinates of the ray $\vec{v}$, known as azimuth ϕ and elevation θ angles, and computed as:

$$\phi = \text{atan2}(v_y, v_x) \tag{A.3}$$

$$\theta = \sin^{-1} \frac{v_z}{\sqrt{v_x^2 + v_y^2 + v_z^2}} \tag{A.4}$$

and are related with the coordinates of the pixels in the panorama though the linear mapping,

$$x_{pan} = x_0 + \frac{\phi}{2\pi}W \tag{A.5}$$

$$y_{pan} = y_0 + \frac{\theta}{\pi}H \tag{A.6}$$

where (x_0, y_0) is the center of the panoramic image, and (W, H) are, respectively, the width and height of the panoramic image in pixels.

Appendix B

Phosphene generation

In order to simulate the visual perception in prosthetic vision, our phosphenes generated are idealized representations of the percepts feasible in the current implants [39, 57–60]. For the generation of phosphenes we start from the final image that is shown by the viewfinder (see Appendix A). First, we calculated the number of phosphenes to represent (200 or 500 phosphenes) depending on image size. Once the size of the phosphenes is known, the position of each phosphene is calculated from the image size, the viewing radius, the size of the phosphenes, the calibration array, and the center point of the image. The result is a array of rays associated with each phosphene. We use this array of rays in real-time, the size of which is much smaller than if we go through the viewer image pixel by pixel.

To generate the phosphenes we use the position and size previously calculated. Different patterns of phosphenes have been used for simulation such as square [61], hexagonal [62] or irregular pattern [44]. One way to specify a pattern configuration is by changing the distance between phosphenes. However, the most commonly phosphene pattern adopted by SPV in the literature is square array [34]. Similarly to many SPV studies [34], phosphenes are approximated as green circular dots with a Gaussian luminance profile. The luminance profile of each phosphene has maximum intensity at the center and gradually decays to the periphery, following an unnormalized Gaussian function $G(x, y)$ defined in Equation (B.1).

$$G(x, y) \propto \exp\left\{\frac{1}{2}\frac{(x - \mu_x)^2 + (y - \mu_y)^2}{\sigma^2 \cdot \mathcal{L}^2(x, y)}\right\} \tag{B.1}$$

where the radial (isotropic) standard deviation of the Gaussian profile is $\sigma \cdot \mathcal{L}(x, y)$ with $\mathcal{L}(x, y)$ being the normalized intensity.

$$\mathcal{L}(x, y) = \frac{\mathcal{I}(x, y)}{\max(l)} \tag{B.2}$$

The intensity of a phosphene is a function $f : \mathbb{R}^n \to \mathbb{R}$ of the intensity of the pixels of the input image, usually the average of the a small window. Each intensity ($\mathcal{I}$) is

quantified to each individual phosphene dynamic range $\mathcal{I}(x,y) \in \{0, \ldots, l-1\}$ where l is the number of intensity levels of the phosphene dynamic range.

Then, the final pixel value for the simulated image is:

$$P(x,y) = \mathcal{L}(x,y) \cdot G(x,y) \tag{B.3}$$

Thus, both the size and brightness are directly proportional to the quantified sampled pixel intensities $\mathcal{L}(x,y)$.

References

[1] Hartong D T, Berson E L and Dryja T P 2006 *The Lancet* **368** 1795–1809
[2] Yu D Y and Cringle S J 2005 *Experimental eye research* **80** 745–751
[3] Zhou D D, Dorn J D and Greenberg R J 2013 The Argus II retinal prosthesis system: An overview *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)* pp 1–6
[4] Cheng D L, Greenberg P B and Borton D A 2017 *Current eye research* **42** 334–347
[5] Lovell N H, Hallum L E, Chen S, Dokos S, Byrnes-Preston P, Green R, Poole-Warren L, Lehmann T and Suaning G J 2007 *Handbook of neural engineering* 337–356
[6] Luo Y H L and Da Cruz L 2016 *Progress in retinal and eye research* **50** 89–107
[7] Zapf M P H, Boon M Y, Matteucci P B, Lovell N H and Suaning G J 2015 *Journal of neural engineering* **12** 036001
[8] Humayun M S, Dorn J D, Da Cruz L, Dagnelie G, Sahel J A, Stanga P E, Cideciyan A V, Duncan J L, Eliott D, Filley E *et al.* 2012 *Ophthalmology* **119** 779–788
[9] da Cruz L, Dorn J D, Humayun M S, Dagnelie G, Handa J, Barale P O, Sahel J A, Stanga P E, Hafezi F, Safran A B *et al.* 2016 *Ophthalmology* **123** 2248–2254
[10] Humayun M 2009 *Investigative ophthalmology & visual science* **50** 4744–4744
[11] Dorn J D, Ahuja A K, Caspi A, Da Cruz L, Dagnelie G, Sahel J A, Greenberg R J, McMahon M J, Argus II Study Group *et al.* 2013 *JAMA ophthalmology* **131** 183–189
[12] Ahuja A K, Dorn J, Caspi A, McMahon M, Dagnelie G, Stanga P, Humayun M, Greenberg R, Argus II Study Group *et al.* 2011 *British Journal of Ophthalmology* **95** 539–543
[13] Ho A C, Humayun M S, Dorn J D, Da Cruz L, Dagnelie G, Handa J, Barale P O, Sahel J A, Stanga P E, Hafezi F *et al.* 2015 *Ophthalmology* **122** 1547–1554
[14] Da Cruz L, Coley B F, Dorn J, Merlini F, Filley E, Christopher P, Chen F K, Wuyyuru V, Sahel J, Stanga P *et al.* 2013 *British Journal of Ophthalmology* **97** 632–636
[15] Luo Y H L, Zhong J J and Da Cruz L 2015 *Graefe's Archive for Clinical and Experimental Ophthalmology* **253** 1907–1914
[16] Dagnelie G, Christopher P, Arditi A, da Cruz L, Duncan J L, Ho A C, Olmos de Koo L C, Sahel J A, Stanga P E, Thumann G *et al.* 2017 *Clinical & experimental ophthalmology* **45** 152–159
[17] Luo Y H, Zhong J J, Clemo M and da Cruz L 2016 *American journal of ophthalmology* **170** 100–109
[18] Kotecha A, Zhong J, Stewart D and da Cruz L 2014 *BMC ophthalmology* **14** 71
[19] Barhorst-Cates E M, Rand K M and Creem-Regehr S H 2016 *PloS one* **11** e0163785
[20] Haymes S, Guest D, Heyes A and Johnston A 1996 *Optometry and vision science: official publication of the American Academy of Optometry* **73** 621–637
[21] Kuyk T, Elliott J L and Fuhr P S 1998 *Optometry and vision science: official publication of the American Academy of Optometry* **75** 174–182
[22] Alshaghthrah A M 2014 *A Study to Develop a New Clinical Measure to Assess Visual Awareness in Tunnel Vision* Ph.D. thesis The University of Manchester (United Kingdom)
[23] Kennedy W L, Rosten J G, Young L M, Ciuffreda K J and Levin M I 1977 *American journal of optometry and physiological optics* **54** 744–755

[24] He Y, Huang N T, Caspi A, Roy A and Montezuma S R 2019 *Translational vision science & technology* **8** 29–29

[25] Ameri H, Ratanapakorn T, Ufer S, Eckhardt H, Humayun M S and Weiland J D 2009 *Journal of neural engineering* **6** 035002

[26] Bloch E, Luo Y and da Cruz L 2019 *Therapeutic advances in ophthalmology* **11** 2515841418817501

[27] Fornos A P, Sommerhalder J, Pittard A, Safran A B and Pelizzone M 2008 *Vision research* **48** 1705–1718

[28] Hayes J S, Yin V T, Piyathaisere D, Weiland J D, Humayun M S and Dagnelie G 2003 *Artificial Organs* **27** 1016–1028

[29] Dagnelie G, Keane P, Narla V, Yang L, Weiland J and Humayun M 2007 *Journal of neural engineering* **4** S92

[30] Vurro M, Crowell A M and Pezaris J S 2014 *Frontiers in human neuroscience* **8** 816

[31] Sanchez-Garcia M, Martinez-Cantin R and Guerrero J J 2020 *PloS one* **15** e0227677

[32] Fornos A P, Sommerhalder J, Rappaz B, Safran A B and Pelizzone M 2005 *Investigative ophthalmology & visual science* **46** 3906–3912

[33] McKone E, Robbins R, He X and Barnes N 2018 *PLoS ONE* **13** e0204361

[34] Chen S C, Suaning G J, Morley J W and Lovell N H 2009 *Vision research* **49** 1493–1506

[35] Xiao J, Ehinger K A, Oliva A and Torralba A 2012 Recognizing scene viewpoint using panoramic place representation *IEEE Conference on Computer Vision and Pattern Recognition* pp 2695–2702

[36] Choi C, Choi M K, Liu S, Kim M S, Park O K, Im C, Kim J, Qin X, Lee G J, Cho K W *et al.* 2017 *Nature Communications* **8** 1664

[37] Haymes S A, Johnston A W and Heyes A D 2002 *Ophthalmic and Physiological Optics* **22** 79–91

[38] Dobelle W H 2000 *ASAIO journal* **46** 3–9

[39] Humayun M S, Weiland J D, Fujii G Y, Greenberg R, Williamson R, Little J, Mech B, Cimmarusti V, Van Boemel G, Dagnelie G *et al.* 2003 *Vision research* **43** 2573–2581

[40] Brelén M E, Duret F, Gérard B, Delbeke J and Veraart C 2005 *Journal of neural engineering* **2** S22

[41] Veraart C, Wanet-Defalque M C, Gérard B, Vanlierde A and Delbeke J 2003 *Artificial organs* **27** 996–1004

[42] Ehrlich J R, Ojeda L V, Wicker D, Day S, Howson A, Lakshminarayanan V and Moroi S E 2017 *American journal of ophthalmology* **176** 26–32

[43] Cha K, Horch K and Normann R A 1992 *Annals of biomedical engineering* **20** 439–449

[44] Cai S, Fu L, Zhang H, Hu G and Liang Z 2006 Prosthetic visual acuity in irregular phosphene arrays under two down-sampling schemes: a simulation study *IEEE Engineering in Medicine and Biology 27th Annual Conference* pp 5223–5226

[45] Chen S, Hallum L, Suaning G and Lovell N 2006 Psychophysics of prosthetic vision: I. visual scanning and visual acuity *International Conference of the IEEE Engineering in Medicine and Biology Society* pp 4400–4403

[46] Stronks H C and Dagnelie G 2014 *Expert review of medical devices* **11** 23–30

[47] Sommerhalder J and Fornos A P 2017 Prospects and limitations of spatial resolution *Artificial Vision* (Springer) pp 29–45

[48] Kihara K and Takeda Y 2010 *Vision research* **50** 2158–2162

[49] Oliva A and Torralba A 2006 *Progress in brain research* **155** 23–36

[50] Oliva A 2005 Gist of the scene *Neurobiology of attention* (Elsevier) pp 251–256

[51] Oliva A and Schyns P G 1997 *Cognitive psychology* **34** 72–107

[52] Denis G, Jouffrais C, Mailhes C and Macé M J 2014 Simulated prosthetic vision: improving text accessibility with retinal prostheses *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* pp 1719–1722

[53] Denis G, Macé M and Jouffrais C 2012 Simulated prosthetic vision: object recognition and localization approach *Proceedings of the 4th International Conference on Neuroprosthetic*

Devices (ICNPD 2012) pp 40–1

[54] Chen S, Hallum L, Lovell N and Suaning G J 2005 *Journal of Neural Engineering* **2** S135

[55] van Rheede J J, Kennard C and Hicks S L 2010 *Journal of vision* **10** 32–32

[56] Vergnieux V, Macé M J M and Jouffrais C 2017 *Artificial organs* **41** 852–861

[57] Humayun M, Yanai D, Greenberg R, Little J, Mech B, Mahadevappa M, Weiland J, Fujii G and de Juan E 2004 Clinical results with the model 1irp implant *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)* vol 1 (IEEE) p 1

[58] Weiland J, Yanai D, Mahadevappa M, Williamson R, Mech B, Fujii G, Little J, Greenberg R, de Juan E and Humayun M 2003 Electrical stimulation of retina in blind humans *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)* vol 3 (IEEE) pp 2021–2022

[59] Mahadevappa M, Weiland J D, Yanai D, Fine I, Greenberg R J and Humayun M S 2005 *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **13** 201–206

[60] Richard G, Hornig R, Keserü M and Feucht M 2007 *Investigative Ophthalmology & Visual Science* **48** 666–666

[61] Chai X, Yu W, Wang J, Zhao Y, Cai C and Ren Q 2007 *Artificial organs* **31** 175–182

[62] Hallum L, Taubman D, Suaning G, Morley J and Lovell N 2003 A filtering approach to artificial vision: a phosphene visual tracking task *Proc. World Congress of Biomedical Engineering and Medical Physics*