Simulating prosthetic vision: Optimizing the information content of a limited visual display

Joram J. van Rheede  
Department of Pharmacology, University of Oxford, UK

Christopher Kennard  
Department of Clinical Neurology, University of Oxford, UK

Stephen L. Hicks  
Department of Clinical Neurology, University of Oxford, UK

Visual prostheses for the restoration of functional vision are currently under development. To guide prosthesis research and allow for an accurate prognosis of functional gain, simulating the experience of a retinal prosthesis in healthy individuals is desirable. Current simulation paradigms lack crucial aspects of the prosthetic experience such as realistic head- and eye-position-dependent image presentation. We developed a simulation paradigm that used a head-mounted camera and eye tracker to lock the simulation to the point of fixation. We evaluated visual acuity, object recognition and manipulation, and wayfinding under simulated prosthetic vision. We explored three ways of optimizing the information content of the prosthetic visual image: Full-Field representation (wide visual angle, low sampling frequency), Region of Interest (ROI; narrow visible angle, high sampling frequency), and Fisheye (high sampling frequency in the center, progressively lower resolution toward the edges). Full-Field representation facilitated visual search and navigation, whereas ROI improved visual acuity. The Fisheye representation, designed to incorporate the benefits of both Full-Field representation and ROI, performed similarly to ROI with subjects unable to capitalize on the peripheral data. The observation that different image representation conditions prove advantageous for different tasks should be taken into account in the process of designing and testing new visual prosthesis prototypes.

Keywords: visual prosthesis, low vision, eye movements, navigation, visual acuity, space and scene perception


Introduction

Loss of vision has a profound impact on quality of life and impairs one’s ability to function independently. Retinitis Pigmentosa (RP) and Age-Related Macular Degeneration (ARMD) are two of the leading causes of non-congenital blindness (RP: Heckenlively, Boughman, & Friedman, 1988; Pagon, 1988; ARMD: Klein, Klein, & Linton, 1992; VanNewkirk et al., 2000; Vingerling et al., 1995). RP affects between 1/3000 and 1/5000 individuals (Pagon, 1988), and ARMD accounts for the majority of permanent loss of vision in people over the age of 50 (Klein et al., 1992; VanNewkirk et al., 2000; Vingerling et al., 1995). These degenerative disorders primarily affect the photoreceptors in the retina while leaving the rest of the retina and visual pathway relatively intact (Kim et al., 2002; Santos et al., 1997; Stone, Barlow, Humayun, de Juan, & Milam, 1992). Electrical stimulation of the remaining retinal circuitry has been explored as a possible strategy for the creation of a visual prosthesis for blind patients, and clinical trials have yielded promising results (Caspi et al., 2009; Humayun et al., 1996, 2003; Rizzo, Wyatt, Loewenstein, Kelly, & Shire, 2003; Yanai et al., 2007) indicating that restoration of at least minimal vision with these implants is feasible. A retinal prosthesis would incorporate an external video camera for image acquisition, an image preprocessor converting the image to a suitable pattern of electrical stimulation, and finally the electrical stimulation array on the retina itself. For an overview of the current state of the art, see, e.g., Chader, Weiland, and Humayun (2009), Dowling (2008), Sachs and Gabel (2004), and Weiland, Liu, and Humayun (2005).

Current concepts for visual prostheses that stimulate the retina vary in size, ranging from 4 × 4 electrode arrays (e.g., Humayun et al., 1999) to over 1000 contact points (e.g., a 1500 electrode array, Zrenner, 2007), but what is clear is that at least at this stage, a retinal prosthesis is going to have only limited resolution. Therefore, any image acquired by a camera will have to be preprocessed and downsampled prior to being converted into a pattern of electrical stimulation.

There are several ways in which such downsampling can be achieved. The most straightforward one is converting the entire captured image to a lower resolution (Full-Field representation). However, other resampling strategies may be adopted, as substantial real-time image processing can be carried out in a wearable device (Tsai, Morley, Suaning, & Lovell, 2009). For instance, it is
possible to zoom in on a certain Region of Interest in the visual scene, increasing the spatial resolution of the visual representation. In this way, a prosthesis wearer will have a greater ability to resolve detail. A drawback of this approach is that it only captures a narrow region of the visual field, leading to loss of peripheral information (i.e., tunnel vision). Another strategy that could be considered to combine the advantages of both is high accuracy in the center of vision combined with low-resolution peripheral information, not unlike fisheye lenses in photography.

Simulating prosthetic vision in healthy volunteers allows researchers to rigorously investigate the minimal requirements for a functional visual prosthesis, to provide an informed estimate of benefit, and to explore the best way to implement such a device. A number of simulation studies have already been conducted, focusing on aspects of visual performance including reading (Cha, Horch, & Normann, 1992; Sommerhalder et al., 2003, 2004), visuomotor tasks (Dagnelie, Walter, & Yang, 2006; Pérez Fornos, Sommerhalder, Pittard, Safran, & Pelizzone, 2008), and mobility performance (Cha, Horch, Normann, & Boman, 1992; Dagnelie et al., 2007), or combinations of these (Hayes et al., 2003; Humayun, 2001; Srivastava, Troyk, & Dagnelie, 2009). A distinct category of simulation studies also aims to provide indications for optimization of prosthesis devices, such as the best way to transfer an image to the stimulation array (Hallum, Suaning, Taubman, & Lovell, 2005) or the optimal configuration of the electrode array itself (Chen, Hallum, Lovell, & Suaning, 2005).

A recent review by Chen, Suaning, Morley, and Lovell (2009) points out that many such simulation studies lack crucial aspects of the prosthetic experience. First of all, subjective reports from clinical trials (Humayun et al., 1996; Richard et al., 2004; Rizzo et al., 2003) make clear that the percept generated by electrical point stimulation of the retina (the “phosphene”) is not like a discrete pixel on a screen but rather resembles a distant point of light with a fading halo, best approximated by a point with a Gaussian luminance profile. Second, the electrode array for stimulation will not move with respect to the eye as it is attached to the retina, so simulation studies should ensure that the simulated phosphene image is gaze fixed. Previous simulation experiments fail to register the image on the subjects’ gaze, allowing subjects to scan the prosthetic image with their eyes, which makes the experimental tasks unrealistically easy. Third, the presentation of the simulated prosthetic vision should be real time and dynamic, i.e., dependent on head (and eye) position as they would be for a patient with a prosthesis (Chen et al., 2009; see also Pérez Fornos, Sommerhalder, Rappaz, Safran, & Pelizzone, 2005 for the functional consequences of offline vs. real-time image presentation and Chen, Hallum, Suaning, & Lovell, 2007 for the functional benefits of a head-movement-dependent image). Other parameters that simulated prosthetic vision studies often fail to incorporate are the range of light intensities that can be discriminated with electrical point stimulation (estimated to be between 8 and 16 gray levels) and the fact that the interface between electrodes and retina is bound to be imperfect leading to irregularities and omissions in the phosphene pattern (Chen et al., 2009).

The aims of this study were to: (1) develop a retinal prosthesis simulation paradigm that is as realistic as possible, taking into account subjective experiences from clinical trials of electrical stimulation of the retina, ensuring accurate phosphene representation, retinal stabilization, and real-time, dynamic image presentation; (2) develop a set of psychophysical tasks for evaluation of visual acuity, object recognition and manipulation, and wayfinding performance with simulated prosthetic vision; and (3) use the simulation paradigm and performance evaluation tasks to investigate different prosthetic vision image preprocessing strategies (Full Field, Region of Interest, and Fisheye).

### Methods

#### Subjects

Twelve healthy volunteers (5 women, 7 men) with normal or corrected-to-normal vision participated in this study. Experiments were conducted in accordance with the ethical guidelines of the University of Oxford, and consent was obtained from the participants before the commencement of the study.

#### Apparatus

The visual scene was acquired using a web camera (Firefly MV IEEE 1394, Point Grey Research) with a resolution of $752 \times 480$ pixels and a frame rate of 60 frames per second. The camera was attached to the head-mounted display and acquired the visual scene in front of the subject (Figure 1). A second camera was used for image acquisition during the wayfinding task, capturing the image on a laptop screen. During the wayfinding task, the built-in motion sensor of the headset was used to create a head-position-dependent image. Horizontal and vertical eye positions were acquired at 250 Hz using a JAZZ-novo eye tracker (Ober Consulting, Poland), which was worn underneath the head-mounted display.

Image acquisition, resampling, and presentation were performed by custom-made software (LabVIEW 8.5 IMAQ, National Instruments) running on a high-end PC to ensure real-time processing. The output image was presented on a Z800 3Dvisor head-mounted display (eMagin) with a resolution of 800 by 600 pixels, spanning a visual angle of 32 degrees horizontally and 24 degrees vertically.
The acquired image was converted to grayscale and resampled to constitute the $30 \times 30$ output image according to the condition (Full Field, Region of Interest (ROI), or Fisheye). For the Full-Field condition, the initial image was cropped to a square and then downsampled to $30 \times 30$ pixels. To improve the sharpness of the scene, downsampling involved specifying 1 pixel per $16 \times 16$ square rather than interpolation. For the ROI condition, a section of the scene was cropped dependent on the subject’s eye position. This square region was then downsampled to $30 \times 30$ pixels. The third condition, Fisheye, was an attempt to combine the high detail of ROI with the peripheral view of the Full Field, thus producing a visual image close to that produced by the fovea/periphery structure of the retina. As a pure Fisheye resampling strategy would lead to a highly distorted image, the central area was “flattened” to a linear sampling frequency equivalent to the ROI sampling frequency. More information of the Fisheye resampling is provided in Appendix A and Figure A1. Examples of each of the image resampling strategies and corresponding images are provided in Figure 2. Real-time examples of these sampling paradigms in action are provided in Movies 1–8.

Sampled pixels were spaced 7 pixels apart to form a regular square array and then displaced randomly by up to 2 pixels in any direction. The final location of pixels was fixed for an entire condition. The resulting image was convolved with a 2D Gaussian function (7 pixels wide) to create the desired halo around the pixels and to allow for partial overlap and fusion of the simulated phosphenes. The resulting image spanned 210 pixels horizontally and vertically, corresponding to approximately 8 degrees of visual angle. To simulate the limited levels of electrical stimulation available in retinal prostheses, the simulated output contained no more than 8 gray levels. Retinal stabilization was achieved by positioning the output image on the current location of the subject’s gaze as acquired by the eye tracker. Real-time drift correction was employed to improve image registration during the tasks, and blinks were detected and used to temporarily blank the visual output. The primary effect of the gaze-based repositioning was to prevent subjects from using eye movements to scan the image.

Subjects were guided through the following sequence of tasks: A benchmark assessment of visual acuity (the
Snellen chart), an object manipulation and visuomotor task (block placement), a face perception task (facial expression judgment), and a set of wayfinding tasks in a virtual 3D environment. Prior to the wayfinding tasks, subjects were familiarized with the 3D environment with normal vision. The tasks were always administered in the same order, but the order of the three image preprocessing conditions (Full Field, Region of Interest, and Fisheye) was randomized to counterbalance practice effects between conditions.

Figure 2. Resampling and image rendering for the different processing strategies. Column 1 demonstrates the 3 different resampling strategies: (B) Full Field, (C) Region of Interest, and (D) Fisheye. Row (A) presents 3 different camera images corresponding to the Snellen task (2), the facial expression task (3), and the block task (4). The images in rows (B)–(D) and columns 2–4 present the image as it is sent to the head-mounted display.
Experimental tasks

Assessment of visual acuity

The ability of participants to resolve visual detail was assessed using the Snellen chart (Figure 3A), a standardized test of visual acuity. The standard test is administered at a distance of 20 ft (where 20/20 vision represents seeing at 20 ft what an average person with normal vision would see at 20 ft), but because of the severely limited vision provided to our participants, the test was administered at 5 ft instead, and the letters were enlarged to 1.5 times their original size, to avoid a floor effect in the Full-Field condition (for a sample trial, see Movie 1). A Snellen chart was obtained from Wikimedia Commons (creator: Jeff Dahl, http://en.wikipedia.org/wiki/File:Snellen_chart.svg) and modified into several different versions (in which the letters were randomly repositioned) and a new chart presented for each condition.

Facial expression judgment

A more continuous assessment of visual acuity was made using a facial expression judgment task (Figure 3B), implemented using Presentation software (Neurobehavioral Systems). Subjects were presented with pairs of faces with different expressions, which could be happy, neutral, or unhappy and asked to press a button on the side of the face that they considered to be the happiest. Faces were presented on a laptop screen with a white background. Trials were self-paced, with an inter-stimulus interval of 500 ms (for sample trials, see Movies 2–4). There were faces of 4 different people, each with 3 different expressions. The three possible expression pairs, therefore, were: happy vs. unhappy, happy vs. neutral, and neutral vs. unhappy. This resulted in 4 (person) × 4 (other person) × 3 (expression pairs) = 48 unique trials. Subjects completed all 48 trials in a randomized order for all 3 image preprocessing conditions. Reaction time (i.e., self-paced trial time) and the percentage of correct responses were determined using data from all 48 trials. Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition, Carnegie Mellon University (http://www.tarrlab.org).

Block task

The head-free nature of our paradigm allowed for the assessment of performance in real visual space. A block placement task was devised, based on the “CHIPs task” in Pérez Fornos et al. (2008), which required subjects to match colored blocks of different shapes with a set of black figure outlines on a workspace area, by placing each block on top of its corresponding outline (Figure 3C). This method allows subjects to combine tactile and visual information when identifying the shapes of the object (for a sample trial, see Movie 5). A number of block sheets were generated for block placement, in which the positions and orientations of the blocks were randomized. Subjects were presented with a randomly selected sheet for each condition.

Subjects completed two block sheets per condition, one of which had only 4 shape outlines and served as a practice trial to get familiar with the task. The other block sheet had 10 outlines and results from the large sheet were used for analysis. Performance was measured by time to completion of the task, the number of blocks correctly matched, and block positioning accuracy. Positioning accuracy was measured by photographing the final block configuration and calculating the ratio between block surface (Figures 3C–4) and the area of the black figures not covered by the blocks (Figure 3C-3).

Wayfinding

One of the most devastating consequences of blindness is that patients lose a great deal of independent mobility. Subjects’ ability to navigate was assessed in a virtual 3D environment, provided by the computer game Half-Life 2 (Valve Software). See Figure 3D for the corresponding landmarks (in brackets): Two tasks were defined for the subjects to carry out:

1. Move through tunnel 1 (1), pass underneath a bridge (3), and return through tunnel 2 (2).
2. Move through tunnel 2 (2), pass underneath a bridge (3), turn to the right and follow the landscape and pass by a large container (4) and then finally turn right again to pass underneath another bridge (5).

Both these paths were strewn with obstacles that subjects had to navigate around (for sample trials, see Movies 6–8). Subjects moved through the environment by means of a joystick (walking backward and forward) combined with the built-in head-movements sensor of the head-mounted display (rotating right and left, looking up and down). The subject’s virtual position was logged at a rate of 1 Hz, and time to completion of the task and path length were recorded.

Analysis

Means were compared by means of repeated measures analysis of variance (ANOVA) procedures, followed by pairwise comparisons between individual conditions. The data were tested for sphericity using Mauchly’s test. In case the assumption of sphericity was violated, the results of Mauchly’s test were reported, and the results of the repeated measures ANOVA were corrected using the Greenhouse–Geisser correction. Significance values for the pairwise comparisons were corrected for multiple comparisons using the Bonferroni correction.

In cases where the measures are not benchmark scores or percentages (i.e., the Snellen score and the percentage of correctly read facial expressions) but arbitrary values
(e.g., time to completion of the block task, navigation time and path length), the task performance is reported relative to performance in the Full-Field condition. An individual’s relative score was therefore computed by dividing performance results in the other two conditions by their result in the Full-Field condition.

Results

Visual acuity

Even at a distance of 1.5 m and at 1.5 times its original size, the Snellen chart proved to be difficult for all subjects (Figure 4A). Ten out of 12 subjects were able to read the top letter in the Full-Field condition, though few could read more than that (mean number of rows: 1, \(SE = 0.19\)). Greater detail could be resolved in both the ROI and Fisheye conditions (mean number of rows for ROI: 3.64, \(SE = 0.24\); Fisheye: 3.54, \(SE = 0.21\)). The effect of the sampling condition was significant \((F(2,20) = 104.23, p < 0.001)\). Pairwise comparisons indicated that the difference between FF and the other two conditions was highly significant \((p < 0.001)\), whereas Region of interest (ROI) and Fisheye conditions were indistinguishable \((p = 1.00)\).

The mean score for FF translated to a Snellen acuity of 20/1200; ROI and Fisheye performances were between 20/420 (row 3) and 20/300 (row 4). A Snellen acuity of 20/200 is the threshold for being legally blind.

Facial expression judgment

The facial expression judgment task provided a behaviorally relevant measure of visual acuity (see Figure 4B, and for an appreciation of the difficulty of the task in the different conditions, see Movies 2–4). Accuracy in the Full-Field condition was narrowly above chance level (mean hit percentage 56.07%, \(SE = 3.22\%)\), whereas the other two conditions allowed for better appreciation of the different expressions (ROI: 77.56%, \(SE = 2.86\%\); Fisheye: 77.08%, \(SE = 2.75\%\)). There was a significant effect of condition \((F(2,22) = 49.52, p < 0.001)\). Pairwise comparisons indicated that the difference between FF and the other two conditions was significant \((p < 0.001)\) while the performances in the other two conditions were indistinguishable \((p = 1.00)\). Figure 4C illustrates that reaction time (i.e., self-paced trial duration) was very similar across conditions (FF: 4.35 s, \(SE = 0.57\) s; ROI: 4.22 s, \(SE = 0.50\) s; Fisheye: 4.35 s, \(SE = 0.53\) s), and any differences were not significant (Sphericity violated: \(\chi^2(2) = 9.01, p = 0.01\); \(F(1.26, 13.81) = 0.43\) with
Greenhouse–Geisser correction, $p = 0.57$), indicating that subjects took equal amounts of time to compare the faces and make a decision across conditions.

### Block task

Most subjects were able to complete the block task with few or no errors, regardless of the condition (Figure 5A).

One subject, however, did not manage to complete the task at all and was excluded from further analysis. The average percentage of correct block matching was 91.8% ($SE = 6.4\%$) for FF, 92.7% ($SE = 3.0\%$) for ROI, and 84.5% ($SE = 4.9\%$) for FE, with no significant differences between any of the conditions ($F(2,20) = 1.00, p = 0.39$).

Task completion was quickest in the Full-Field condition (Figure 5B, Movie 5), with an average time of 129 s ($SE = 14$ s). Completion times for ROI and Fisheye were

---

**Figure 3.** Experimental tasks. (A) A Snellen chart used to determine an ophthalmological benchmark score of visual acuity. (B) A sample trial of the facial expression task: Participants were asked to select the happier of two simultaneously presented faces. (C) The block task. Participants were given a white sheet with black shapes (1) corresponding to a set of colored blocks. They were asked to place the blocks on top of the shapes, as shown in (2). To determine the accuracy of block placement, the ratio between uncovered black shapes (3) and block surface (4) was determined. (D) The wayfinding task. Participants were asked to carry out 2 navigation tasks in this virtual 3D environment, here presented from a bird’s-eye view. For a description of both tasks, see main text.
177 s ($SE = 18$ s) and 181 s ($SE = 24$ s), respectively. Relative to Full-Field performance, subjects took 1.40 times as long ($SE = 0.11$) to complete the task in the Fisheye condition and 1.53 times as long ($SE = 0.23$) in the ROI condition. However, the effect of condition was not found to be significant (Sphericity violated: $\chi^2(2) = 6.37$, $p = 0.04$; $F(1.33, 13.27) = 2.904$ with Greenhouse–Geisser correction, $p = 0.104$) even though pairwise comparisons hinted at a significant difference between Full-Field and Fisheye performances ($p = 0.02$).

An increase in positioning accuracy was observed for the ROI condition but not for the Fisheye (Figure 5C). The average error score for ROI was 24.3 ($SE = 2.2$), while Full-Field and Fisheye conditions maintained error scores of 36.2 ($SE = 4.5$) and 32.6 ($SE = 3.9$), respectively.

The effect of condition on positioning accuracy was found to be significant ($F(2,20) = 4.11$, $p = 0.032$). ROI error score was 0.73 times that of Full Field ($SE = 0.07$), a difference found to be significant ($p = 0.012$) using pairwise comparisons. The difference between ROI and Fisheye error score was not significant ($p = 0.18$), even though Fisheye error score was nearly equivalent to Full Field, with a relative score of 1.00 ($SE = 0.13$, $p = 1.00$).

**Wayfinding**

One subject complained of nausea and was unable to complete the wayfinding task. For the other subjects, the Full-Field representation appeared to be the most favorable for wayfinding purposes (Figures 6A and 6B). First, sampling condition had a significant effect on wayfinding time for task 1 ($F(2,20) = 6.49$, $p = 0.007$). On average, subjects required 68 s ($SE = 12$ s) to perform task 1,
whereas they needed 1.84 ($SE = 0.24$) times as long in the ROI ($p = 0.001$) and 1.62 times as much ($SE = 0.24$) in the Fisheye condition (not significant, $p = 0.14$). There was no significant difference between performances using Fisheye and ROI ($p = 1.00$). On task 2, a similar pattern was observed, though the overall effect of condition was not found to be significant ($F(2,20) = 3.11, p = 0.066$). Subjects took an average of 78 s in the FF condition ($SE = 15$ s), 1.59 times as long ($SE = 0.37$) for ROI, and 1.58 times as much ($SE = 0.19$) in the Fisheye condition. Interestingly, pairwise comparisons hinted at a difference between FF and Fisheye task 2 wayfinding times ($p = 0.009$).

To investigate whether the longer time course for the ROI and Fisheye conditions was due to either slower progress or a greater number of detours, the length of the paths taken by the subjects was compared across the different conditions. In task 1, a significant effect of condition on wayfinding path length was found ($F(2,20) = 10.86, p = 0.001$). Subjects in the ROI condition covered on average 1.37 times ($SE = 0.08$) the amount of distance

Figure 6. Wayfinding task performance. (A) Time to completion of wayfinding tasks 1 and 2, relative to Full-Field performance. (B) Length of the path traveled during wayfinding tasks 1 and 2, relative to Full-Field performance. (C) Paths of individual participants during the wayfinding task, for tasks 1 and 2, for each of the resampling strategies (FF, ROI, and FE). Error bars represent standard error of the mean. *$p < 0.05$; **$p < 0.005$; ***$p < 0.001$. 

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933538/ on 04/06/2018
covered in FF ($p = 0.002$). In the Fisheye condition, subjects covered 1.35 times ($SE = 0.09$) as much distance ($p = 0.003$). Fisheye and ROI did not differ significantly ($p = 1.00$). In task 2, a significant effect of condition was also found ($F(2,20) = 5.28, p = 0.014$). Subjects in the ROI condition covered 1.20 times ($SE = 0.09$) as much ground as in the FF condition (not significant, $p = 0.11$) and 1.31 times as much ($SE = 0.09$) in the Fisheye condition ($p = 0.019$). The difference between Fisheye and ROI was not significant ($p = 0.79$).

### Practice effects

It is difficult to approach in an experiment the amount of training time an actual prosthesis wearer has (though in an ambitious study by Pérez Fornos et al. (2008) subjects were trained for more than 15 sessions). In our study, subjects were not trained prior to participating in the experiment, except for being guided through the virtual environment once for each navigation task. While the Snellen chart and the facial expression judgment task rely mainly on visual acuity, there is a serious potential for practice effects inherent to the block task and the navigation task. As subjects proceed through the experiments, they can adopt certain strategies to increase speed and accuracy and may start adapting to the setup. While the order of conditions was counterbalanced across subjects to even out practice effects, it is still interesting to investigate whether such effects are present in our data. This will also provide an indication of how much variability these practice effects have caused (increasing the margins of error for the comparisons).

Block task time to completion, block task accuracy, and navigation task time to completion (for task 1) were investigated to see if practice effects were present. An overview of practice effects is given in Figure 7. Though performance on all tasks seems to follow a trend of improvement, none of the trends reached significance.

### Discussion

This study successfully developed and tested a realistic paradigm for the simulation of a retinal prosthesis. Phosphenes, the percepts generated by electrical point stimulation of the retina, were successfully modeled with a Gaussian luminance profile and irregular spatial distortions inherent to the interface between an electrode array and the retina. The limited ability to discriminate electrical stimulation levels was simulated by restricting phosphene appearance to only 8 gray levels.

The presentation of the phosphene image was real time and dynamic, i.e., head and eye position dependent. Head position dependence was assured by positioning the scene camera on the head-mounted display, or in the navigation task, by using the motion sensor in the head-mounted display to move a virtual camera. Eye position dependence was implemented by means of an eye tracker and was used both for determining the focal point of the image (in ROI and Fisheye conditions) and for ensuring retinal stabilization. Although gaze-fixed stimuli were difficult to maintain, the eye-position-dependent placement of the image ensured that subjects were unable to scan the phosphene array—a distinct improvement on previous prosthetic simulation studies (overview in Chen et al., 2009).

In this study, three different image processing techniques were tested in visual and behavioral tasks: Full Field, Region of Interest, and Fisheye. The motivation behind the Fisheye resampling routine was to combine low-resolution peripheral visual information with a detailed...
central image resulting in a prosthetic visual field with a high-acuity linearly sampled center region, surrounded by a progressively downsampled periphery.

The results of the visual acuity tasks (i.e., the Snellen chart and the facial expression judgment task) clearly show the advantage of ROI over Full Field. Moreover, these tasks illustrate that Fisheye is equivalent to ROI when it comes to ability to resolve detail. This is not unexpected, as sampling frequency of the central (flat top) area of the Fisheye was equivalent to the sampling of the ROI. Nevertheless, this task provides a good baseline and illustrates that Fisheye has at least partially the same advantages as ROI.

The block task provides a clear dissociation of the advantages and disadvantages of Full-Field vs. ROI vision. There was a trend suggesting that Full-Field vision allows subjects to complete the task in less time, presumably because of its greater scene overview allowing for easier location of blocks and one’s own hands. ROI vision, on the other hand, allowed for greater accuracy. Unfortunately, Fisheye performance, rather than incorporating the best of both approaches, was as slow as ROI and as inaccurate as Full Field. The absence of a Fisheye advantage for visual search may be due to the distortion in the periphery, which did not allow subjects to correctly identify shapes in this area. The lack of accuracy could be due to the restricted size of the “flat top” area with linear sampling. Subjects could not fit an entire block in their linear sampling area, necessarily resulting in distortion near the edges.

For the purpose of wayfinding, a Full-Field representation proved to be advantageous. Compared to the ROI and Fisheye conditions, subjects were faster and more efficient in getting from the starting point to the target. ROI and Fisheye performance were essentially equivalent. Subjects appeared to be unable to capitalize on the peripheral data provided by the Fisheye. This may be due to the different velocities of optic flow that a spherical lens produces. It is possible that practice may improve the subjects’ visual perception. A brief exploration of practice effects did not point to any significant effect of practice within our experiment, but it is highly likely that future prosthetic wearers will improve their visual performance over time.

The Fisheye process that we employed was not able to combine the best aspects of Full-Field and ROI representations. Indeed, a form of vision as complex and non-intuitive as the Fisheye may require more training before its full potential can be effectively utilized.

While visual prosthesis research is advancing rapidly in terms of microelectronics and implantation techniques, the issue of what type of visual information will be delivered through the prosthetic display deserves more attention. Importantly, our results highlight that different types of representation of the visual scene in simulated prosthetic vision prove advantageous for the performance of different tasks. This should be taken into consideration during the development of visual prostheses for clinical trials and application.

An experimental approach such as the one described here, for instance, can be used for the future evaluation of more elaborate image processing approaches. For example, it was shown in the current experiment that Full-Field and ROI representations each have distinct advantages. These are not, however, mutually exclusive. Performance using a subject-controlled zoom level, where minimal zoom level would correspond to the current Full Field and a maximum zoom level could correspond to the current ROI, can also be investigated. Further considerations to enhance the visual prosthesis experience might incorporate edge detection algorithms and contrast adjustment to maximize the information that can be obtained from the image.

**Conclusion**

The current study has successfully set up a simulation paradigm for the systematic investigation of questions relevant to the development of retinal prostheses and their expected benefit. The parameters for the simulation of the prosthetic vision were designed to approximate the type of vision reported clinically from retinal prostheses; however, the effects of resolution and optimization may be instructive for other visual prosthesis efforts, such as optic nerve (e.g., Delbeke, Oozeer, & Veraart, 2003) and cortical (e.g., Dobelle, 2000) stimulation.

The set of tasks designed to evaluate visual performance managed to dissociate different types of prosthetic vision successfully. A Full-Field and a zoomed (Region of Interest) representation of the visual world were compared with a Fisheye representation designed to incorporate the strengths of both. The Fisheye was not superior to either of these two conditions, although the effect of training on the differences between image representation conditions is yet to be investigated. Importantly, this study has reported that different image processing techniques prove advantageous for different tasks, which should be taken into account in the process of designing and testing new visual prosthesis prototypes. The retinal prosthesis simulation paradigm and the set of psychophysical tasks developed in this study can easily be used or adapted to answer further questions concerning prosthetic vision to inform retinal prosthesis developers and prospective recipients alike.

**Appendix A**

**Supplementary material**

**Fisheye resampling**

The Fisheye resampling paradigm aimed to combine the advantages of high-resolution tunnel vision (ROI) and
Figure A1. The Fisheye resampling paradigm. X and Y indices in an image, increasing in a linear fashion. (A) The X coordinates of pixels in an image. (B) The Y coordinates, ranging from −50% to +50% where 0 represents the center point of the image. They are multiplied by a function (C) that represents the distance from a central point, where 0 is the minimum and 1 is the maximum distance. Now the sampling increments for the X and Y indices can be visualized as in (D) and (E), sharply increasing around the edges for coarser sampling of the periphery of the image, and more slowly increasing in the center, for a higher sampling rate and therefore greater visual acuity at the center of gaze.
low-resolution overview vision (Full Field). A central part of the resampled image contained a relatively high-resolution sampling frequency, whereas the periphery was represented less accurately as a function of the distance from the center (see Figure 2 for the result).

Whereas the Full-Field and ROI paradigms sample the original image in a linear way (visualized as the planes in Figures A1A and A1B), the Fisheye sampling was produced by multiplying the linear sampling matrices by a distance function that had a value of 0 at the center of vision and increased to 1 in the furthest periphery (Figure A1C). The resulting sampling matrices can be visualized as the curved planes in Figures A1D and A1E. Finally, to avoid excessive distortion in the center of vision, the sampling was made linear in the most central area.

Acknowledgments

This research was supported by the Wellcome Trust and the NIHR Biomedical Research Centre, Oxford.

Commercial relationships: none.

Corresponding authors: Joram J. van Rheede and Stephen L. Hicks.

Email: joramvanrheede@gmail.com; stephen.hicks@cneuro.ox.ac.uk.

Address: Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK; Department of Clinical Neurology, University of Oxford, Level 6, West Wing, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK.

References


