A scalable data transmission scheme for implantable optogenetic visual prostheses

Zhenyang Hou, Walid Al-Atabany, Ramy Farag, Quoc C Vuong, Andrey Mokhov and Patrick Degenaar

1 Newcastle University – School of Engineering, Newcastle upon Tyne, NE1 7RU, United Kingdom
2 Department of Biomedical Engineering, Helwan University, Cairo, Egypt
3 Department of Mechatronics, Helwan University, Cairo, Egypt
4 Newcastle University – Biosciences Institute, Newcastle upon Tyne, NE1 7RU, United Kingdom

E-mail: patrick.degenaar@newcastle.ac.uk

Abstract

Objective. This work described a video information processing scheme for optogenetic forms of visual cortical prosthetics. Approach. The architecture is designed to perform a processing sequence: Initially simplifying the scene, followed by a pragmatic visual encoding scheme which assumes that initially optical stimulation will be stimulating bulk neural tissue rather than driving individual phosphenes. We demonstrate an optical encoder, combined with what we called a zero-run length encoding (zRLE) video compression and decompression scheme—to wirelessly transfer information to an implantable unit in an efficient manner. In the final step, we have incorporated an even power distribution driver to prevent excessive power fluctuations in the optogenetic driving. Significance. The key novelty in this work centres on the completeness of the scheme, the new zRLE compression algorithm and our even power distributor. Main results. Furthermore, although the paper focusses on the algorithm, we confirm that it can be implemented on real time portable processing hardware which we will use for our visual prosthetics.

1. Introduction

The visual system is arguably the most important sense for human beings. Its loss is, therefore, a significant burden on both the individual and their families. It is estimated by the World Health Organization (2014 report [1]) that around 39 million people across the world are legally blind. The definition for this is: ‘Having less than 20/200 vision, or a visual field of fewer than 20 degrees in the better eye’. As such, most of this cohort will have some preserved perception but with visual field deficits. More serious are those with complete visual loss in both eyes through disease or trauma.

Thankfully, in recent years there has been significant pharmaceutical progress to slow the progression of conditions such as wet age-related macular degeneration and diabetic retinopathy. Furthermore, gene therapies are also becoming available for certain forms of hereditary disorders, such as choroideremia [2]. However, such approaches cannot restore function where vision is lost due to degeneration of, or trauma to, the retina. There has been a concerted effort spanning many decades to restore sight via neuroprosthetic means. These can be characterized by the location in the visual system to which stimulus is applied: (i) the retina, (ii) the optic nerve (iii) the lateral geniculate nucleus (LGN), or (iv) the visual cortex. Each approach has its strengths and weaknesses in its efficacy, invasiveness and complexity. As a result, their target conditions for treatment are different. For those conditions in which the communication cells of the retina are still intact, retinal prosthesis is the least invasive and most appropriate route. For conditions in which there is no longer a communication link between the retina and the brain, LGN or visual cortical prosthesis are the most appropriate routes.

The primary conditions applicable to retinal prosthesis is retinitis pigmentosa (prevalence (~1:3000
[3, 4]). In this condition, the light-sensing cells in the eye slowly degenerate from the periphery to the centre, resulting in a loss of functional vision sometime in middle age [5]. In 1992, Stone and colleagues [6] discovered that only the photoreceptors were destroyed in the blinding disease Retinitis Pigmentosa. This discovery opened the possibility of visual restoration by stimulating the remaining retinal layers, and thus a less invasive implant than cortical or LGN approaches. Retinal prosthetics subsequently became the focus of the field for many years [7–9]. Notable clinical implementations include the Argus II [10], and alpha IMS [11] implants, and more recently, the IRIS implant [12]. However, the applicable population is very small and has resulted in economic failure within the retinal prosthetic domain to date, despite clinical approvals. As such, a more recent approach, demonstrated by Mathieson et al [13], is to target a subset of dry age-related macular degeneration (global prevalence: 8% [14]) with geographic atrophy. In this case, individuals have normal vision in most of their retina, but have lost function in the high acuity macular region. It is therefore hoped that with sufficiently high-resolution retinal prosthetics, useful high acuity function can be restored.

For visual cortical (and LGN) prosthesis, the primary applicable conditions include bilateral ocular trauma, bilateral retinoblastoma and glaucoma (prevalence ~1:100 [4, 15]). The latter condition is the most prevalent of these and is related to intracellular pressure causing degeneration of the optic nerve. Despite various available treatments, around 3.3% of those with glaucoma are rendered fully blind [16]. Arguably the first exploratory experiments were performed in 1929 under Forster [17] who explored percepts through direct electrical stimulation of the visual cortex. The first to demonstrate the efficacy of an implantable visual prosthesis was Brindley and Lewin in 1968 [18], followed by Dobelle et al in the 1970s [19]. Recent economic and technical difficulties with retinal prosthetics have renewed interest in brain prosthetics for which there may be a larger applicable patient cohort. For example, prosthetics targeting the visual part of the thalamus (the LGN) has been explored by Vurro et al (2014) [20] has explored the efficacy of prosthetics for the visual part of the thalamus (The lateral Geniculate Nucleus or LGN). Similarly, there has also been both academic [21, 22] and industrial interest in visual cortical prosthesis. At the time of writing, the SecondSight company have embarked upon clinical trials of their Orion implant (NIH Clinical trials ref: NCT03344848).

Information undergoes initial spatio-temporal processing in the retina, separates into parvocellular, magnocellular and koniocellular visual pathways. It is then transmitted to the visual cortex via the LGN. The first entry point into the visual cortex (see figure 1) is V1 (also referred to as Brodmann’s area 17 or striate cortex). V1 is surrounded by V2 and V3, and together these early visual areas process simple visual features such as contrast, edges and orientation, and colour and motion. These visual areas have a retinotopic organisation; that is, adjacent points in visual space map to adjacent positions on visual cortex [23–25]. For V1 in particular, there is also evidence for cortical magnification, with more physical cortex for foveal compared to peripheral vision leading to higher spatial resolution in the fovea [18]. The foveal vision is represented predominantly towards posterior portions of V1, whereas peripheral vision is represented in more anterior portions. V1 also projects to V4, which further processes colour, and V5/MT, which further processes motion. For these reasons, the V1 is the most obvious target for a visual cortical prosthesis.

Due to the different nature of any stimulation array relative to natural vision, an important issue for cortical prosthesis is plasticity in adults. Broadly defined, cortical plasticity describes the cortex’s ability to change its structure or function in response to experience (for a review, see Beyeler et al [26]). These changes can occur at different spatial and temporal scales across the cortex, including the different visual areas discussed previously. In adults, studies which train adults on visual tasks can change their cortical responses in visual areas. For example, researchers have shown enhanced BOLD responses in V1, as measured by functional magnetic resonance imaging, to trained relative to untrained orientations in various tasks (e.g. [27, 28]). There is mixed evidence whether there are changes to the retinotopic organisation of visual areas (see Beyeler [26] for a discussion).

An anatomical-surgical challenge stems from the fact that the visual cortex is found in the medial plane—i.e. where the two hemispheres face each other. Functional imaging studies by Schwarzkopf et al [29] and Dougherty et al [30] suggest that the functional surface area of V1 in each hemisphere is approximately 1500 mm² (~900 mm² to ~2300 mm²), depending on the individual. However, inserting devices between the medial surface and the Falc Cerebri becomes increasingly difficult with depth, limiting the potential field of view in the superior quadrants. Furthermore, a significant portion of the V1 lies within the calcarine sulcus, which is even more inaccessible. Studies by From Srivastava et al [31] showed that stimulus of the accessible V1 gyrii would yield phosphenes in the top and bottom quadrants of the visual field. This would produce an ‘hourglass’ type vision. This was indeed seen when Brindley and Lewin inserted a planar array of equidistant surface electrodes into the human visual cortex. The map of perceived phosphenes followed an hour-glass formation in each hemisphere. A greater field of view was observed in the inferior compared to the superior plane as the latter was deeper into the fissure between the two hemispheres of the brain and thus less accessible. This is corroborated by preclinical studies by others [23–25]. Nevertheless, it may be possible to
Figure 1. Concept of information transfer in a visual prosthesis. (a) A visual prosthesis would utilise a headset to gather information from the visual scene which is then processed by an external video encoder which then wirelessly transfer information to the internal control unit. (b) The internal control unit transfers information to individual optrode arrays which optically stimulate the target tissue. However, optrode arrays can only be place on the gyrus rather than the sulci. (c) Medial view of a brain hemisphere with V1 to V3 of the visual cortex. (d) Optrodes will form a pattern of ‘phosphenes’ which increase in size with eccentricity. It could be challenging to get all the angles from V1 due to parts being in the sulci, but it may be possible to attain mapped areas in V2 and V3. Figure 1 artwork created by Matt Briggs at clinicalillustration.com.

achieve lateral quadrant phosphenes via stimulus of the V2, V3. This is illustrated in figure 1, where red spheres represent a distribution of stimuli in the surface areas of V1. The visual field is smaller in the superior to inferior domains to represent accessibility issues. Green and blue spheres represent conceptual effects from stimulation of V2 and V3.

Most trials to date have utilised electrical stimulation of the nervous system. However, there may be some significant advantage in the use of optogenetics—the genetic photosensitization of nervous tissue. Using gene therapy to target specific types of neurons to be sensitive to particular optical wavelengths has significant potential in neuroprosthetics as a whole. We have previously explored optogenetic approaches to retinal prosthesis [32, 33] and Luo et al [34] recently demonstrated how are technology could be used for closed-loop control methodologies. A similar approach could potentially be used to improve visual prosthetics.

This paper considers these issues in detail to present a full processing stream for the Newcastle Visual Cortical Prosthesis, whose concept is presented in figure 1. In particular, we present a compression and mapping scheme for the transmission of information to each optrode. Furthermore, we present an even power distribution scheme to ensure maximum use of the individual stimulators.

2. System design rationale

The overall concept of a visual cortical prosthesis is illustrated in figure 1. A headset with a built-in imaging system acquires the visual scene. The video frames is processed in real-time by an external computer to extract and compress relevant visual information, which is then transmitted wirelessly to an implantable control unit. This unit then transmits a signal to individual optrode units to stimulate the visual cortex. In order to design the video processing sequence, we consider four key architectural constraints: (i) spatial resolution, (ii) contrast, (iii) temporal resolution, and (iv) thermal management.

An important aspect with regards to our approach is that we utilise a relatively simplistic visual cortical encoder, i.e. we mimic the bypassed functions from the retina to V1 as simple spatial and temporal derivatives. We nevertheless recognise other important work in this field. For example, Nirenberg and Chetham [35] presented an encoder for optogenetic retinal prostheses in which they postulated a very high level of visual return could be restored. Similarly, Jepson et al [36] have explored models to reproduce the spatio-temporal signalling of the retina for retinal prostheses. More specific to the visual cortex, Martinez et al [29] have described a model of the information transmission across simple and complex cells. These approaches, including ours, will become increasingly important as the fidelity of the prosthetic system shifts from stimulating neural tissue towards individual cells, and it becomes possible to identify cells being stimulated.

Finally, it should be noted that when we transfer stimulation to the internal control unit, we assume these to be in the form on pixel intensities and thus image encoded rather than neural encoded. In the reverse case, it may well be in the future that we would want to extract neural data from the stimulated visual cortex. In this case, neural compression regimes such as that presented by Sun et al [30] could be used. However, that is for future work.
2.1. Spatial resolution considerations

Dagnelie et al [37] demonstrated that even 256 phosphene percepts were sufficient for basic reading tasks. Other past efforts at simulating visual stimuli have been presented by Chen et al [38, 39], Lewis et al [40] and Yue et al [41], with similar results. To date, in the visual cortical prostheses domain, the original Brindley system had 80 electrodes (40 in each hemisphere) [18] and the subsequent Dobelle system 64 [19]. The UMH team have explored trials using the Utah array with up to 100 stimulation sites [42, 43], and more recently the Second Sight company are trialling an Orion implant (Clinical trials ref: NCT03344848) with 60 stimulation sites. Finally, the Illinois and Monash systems aim for hundreds of implanted electrodes with distributed 16 electrode brain units (as presented in the Eye and the Chip Conferences)—but these have not yet gone to trial. In the retinal prosthesis domain, the Argus II retinal prosthesis has had 60 stimulation sites [10], and The alpha IMS system has 1500 stimulation sites [11], though the latter case did not typically translate into better vision than the former. For optogenetic retinal prosthesis, we have previously proposed an 8100 (non-implantable head-set) optical stimulation system [33], but this has not undergone clinical trials. However, to date, prior clinical efforts have not produced a truly transformative visual return.

Table 1 describes an exemplar approach of how the number of stimulating sites could be scaled according to the architecture presented in figure 1. We assume that the primary target for stimulation are simple cells in layer 4 of the V1 portion of the visual cortex [29]. As such, at this stage, we do not consider multiple LEDs per optrode shaft. W reflect on our own prior studies [44–46] and the clinical efforts to date with hundreds of stimulators that have produced disappointing levels of vision. As such, we estimate that a target in excess of 1000 independently discernible stimulation sites would be required to bring back functional vision. Thus for this study, we consider the numbers of stimulators around this value: [512, 1024, 2048]

2.2. Contrast considerations

We have noticed in prior studies [45] with low vision patients that contrast sensitivity is as important as visual acuity in predicting how well these patients perform in visual tasks. This needs to be considered for the case of optogenetic forms of visual prosthetics. The response of channelrhodopsin-2 encoded neurons to light is often given in terms of an irradiance threshold of 0.7 mW mm\(^{-2}\) [47], and similar for other opsins variants. However, the neural response is perhaps better described as an S-curve of neural response to the logarithmic stimulus of light as per [33]:

\[
R = \frac{\beta}{\beta - 1 + e^{-\alpha \cdot \log(\Phi_{\lambda})}} = \frac{\beta}{\beta - 1 + \Phi_{\lambda}^{-\alpha}} \quad (1)
\]

where refers \(R\) to the normalised neural response (to the normalised radiance), \(\alpha\) is a slope factor and \(\beta\) defines the slope centre. \(\Phi_{\lambda}\) is the normalised radiance—which can be presented as either a short pulse width modulation, or as a pulse amplitude modulation. Given short frame times, it can be more convenient to do the latter as per our past LED driving schemes [48–50]. \(\Phi_{\lambda}\) is therefore quantized between 0 and 255 to provide a full dynamic range. It can, therefore, be defined as:

\[
\Phi_{\lambda} = \left(\frac{R' (1 - \beta) + \beta}{R'}\right)^{-\frac{1}{\beta}} \quad (2)
\]

Where \(R'\) is the required neural response, defined by the video processing algorithms. The response is normalized and quantized according to the bit resolution (typically 8 bits for most video processing). It should be noted, that the upper saturation of the response curve requires significant irradiance—typically around \(10^3\) mW mm\(^{-2}\) on the cell. In contrast, the lower saturation limit can be as low as \(10^{-2}\) as per Barrett et al [51].

2.3. Temporal resolution considerations

Similar to spatial resolution, temporal vision can be characterized by a temporal contrast sensitivity function which measures temporal acuity for different contrast modulation frequencies as a function of luminance contrast. Several important low-level visual processes including the detection of contrast, edge and texture; binocular-disparity changes (flicker); detection of first-order motion; and binding of local elements into global form have a fast temporal resolution on the order of 20 Hz to 50 Hz (Holcombe [52]). As such, ‘video rate’ of how the scene is changing would need to be presented to the user at least 25 frames per second. For comparison, the target for consumable virtual and augmented reality is to present video between 50–100 Hz. Also, it should be noted that the latency between image capture and presentation should be minimal, preferably not more than a single frame i.e. less than ~40 ms.

Our scheme currently takes a naive assumption that the user’s eye is fixed in space and that the camera on the headset moves smoothly with the head. In reality, this ignores the phenomenon of saccadic suppression of fast eye movements [53]. Furthermore, we can expect motion artefacts to occur with vibrations and movement of the headset if there is no built-in stabilization mimicking the vestibulo-ocular reflex.

2.4. Thermal management

The key caveat to the optogenetic technique is that optically sensitized cells require considerable irradiance to be activated by light. That light needs to be created from a source which is typically significantly less than 100% efficient. As such, heating can become a concern. We, therefore, follow the regulatory guidance set by the American Association of Medical
### Table 1. Target optrode resolutions for an optogenetic clinical visual prosthesis using the architecture described in figure 1.

| Optrodes per brain implant unit | Double-sided emission? | Brain units per hemisphere | Hemispheres | Total |
|-------------------------------|-------------------------|---------------------------|-------------|-------|
| 16                            | N                       | 4                         | 2           | 128   |
| 16                            | N                       | 8                         | 2           | 256   |
| 16                            | N                       | 16                        | 2           | 512   |
| 32                            | N                       | 16                        | 2           | 1024  |
| 32                            | Y                       | 16                        | 2           | 2048  |

Instrumentation (AAMI), which recommends a limit of $\Delta T = +2^\circ C$. Prior modelling effort has determined the architecture and efficiency limits which this sets on individual optrodes [54]. From a processing perspective, we need to ensure that our optical stimulation is at the maximum possible efficiency to prevent thermal harm. There is a further consideration in that individual light sources can consume mA of current for short periods of time. However, the maximum current that can be delivered will be in the tens of mA. As such, there needs to be an interleaving scheme between LED operations.

### 3. System architecture

The processing scheme is described by figure 2. We propose a series of steps to optimally process the image on both the external (headset) and internal (subcutaneous control) units with which to provide required visual information in an optimised way. These are described in the following sub-sections.

#### 3.1. Acquisition

We have previously explored multispectral imaging approaches [55], but for this work, we propose a typical video stream from a standard portable camera with automatic gain control. The camera will be most likely located on a headset from which information can be passed to a video processor. The challenge with this approach is that the eye and the camera would move independently from each other. That, in turn, may inhibit the ability of the visual cortex to integrate data into a general understanding of the scene. However, that is for future work to explore. Once the visual information is passed to the video processor, the image is resized to the required resolution according to table 2, and converted to greyscale. Then it is passed on to the following processing stage.

#### 3.2. Pyramidal (log-polar) image resizing

In typical imaging and display systems, we acquire and display images in a uniform grid of equally sized pixels. However, when Brindley and Lewin [18] implanted a planar electrode array onto the (medial surface of each hemisphere) visual cortex, they found that perceived phosphes at a large eccentricity of 35° were four times larger than those perceived at the centre (i.e. 0° eccentricity). This is in keeping with the foveated nature of primate vision—which perceives higher resolution (denser pixels) at the centre and lower resolution (less dense pixels) in the periphery.

One mathematical method to describe foveated scenes is the log-polar model. In this case, the location and size of pixels are decided by the formula $r = a^n/k$, where $a$ and $k$ are constants deciding the number and arrangement of pixels, $r$ is the radius of $n^{th}$ layer of pixels, and the size of phosphes can be calculated via $Sz = (r_i - r_{i-1})$. Unfortunately, mapping a cartesian image to this form of log-polar arrangement makes subsequent image processing challenging, i.e. computationally inefficient. Alternatively, if the image processing is performed in the cartesian image, at a resolution equal to the maximum resolution of the foveated centre, then it would also be computationally wasteful.

As such, we utilize a specific adaptation of the pyramidal scaling approach as defined by Andelson et al [56], albeit for the purposes of foveation rather than convolutional scaling. The basic principle is to have three levels as per figure 3(b)—each with an areal resolution a quarter of the previous level. Specifically, we take the following steps: we (i) resize the image to the target processing size; (ii) extract the central third of the image (with a couple of padding pixels) to create a segment representing the high-resolution fovea/macula; (iii) rescale the image by half and cut out the central two-thirds segment to act as the mid resolution layer, and finally (iv) rescale down again by half again to attain the low-resolution periphery. The layers can then be processed separately as per section 3.3 and then mapped to a phosphene/stimulation arrangement as per section 3.4.

Details on layer and segment size for the pyramidal approach can be seen in table 2. The advantages are twofold: filters do not need to increase in size for the periphery, and the total number of pixels to be processed is reduced by 66%. As such, the overall effect is to reduce the processing by an order of magnitude, thus allowing for implementation on portable processing platforms.

The outcome of the pyramidal foveation approach can be seen in figure 3. (a) Shows an original image to be processed. (b) Shows the pyramidal layers which have been flattened in (c) to show how the image scales in resolution from centre to periphery. (d) Shows the image simplification and retinal processing as per section 3.3. (e) Shows a simulated ‘perfect’ arrangement of phosphes if they followed the log-polar principle described previously. (f) Shows...
the outcome (using the higher $192 \times 192$ resolution) if the image was directly mapped directly to the phosphene arrangement in (e) to the log-polar image. (g) Shows the outcome if the pyramidal version of the image was directly mapped to the log-polar image.

### 3.3. Video processing

Despite our best efforts, the visual return for the foreseeable future will be poor compared to normal vision. As such, it is important to maximise the useful information to be presented to the user. In a previous effort on low vision patients with age-related macular degeneration, we found that enhancing contrast, in particular, has some beneficial effect \[45\]. We utilise an optimised version of image processing previously described in [44], and summarised in figure 1. Namely we (i) perform simplification of the visual scheme, and (ii) perform retinal processing—both on each of the pyramidal layers prior (section 3.2) to phosphene mapping (section 3.4). A more detailed sequence of these processing steps can be seen in the supplementary information (available online at stacks.iop.org/JNE/17/055001/mmedia). In summary, we perform the image simplification processing using an anisotropic algorithm:

\[
I^{n+1} = I^n + \Delta t \left[ \nabla C \cdot \nabla I_H + \nabla C \cdot \nabla I_V \right] \tag{3}
\]

Where $I$ is the initial unprocessed image, $C$ is the diffusion function, which monotonically decreases as a function of the image gradient value. $\nabla$ represents the gradient operator, $\Delta t$ is the time step (controls the smoothness speed), and $n$ is the iteration/recursion number. $\nabla I_H$ and $\nabla I_V$ represent the gradients of the image. The degree of smoothing is determined by the number of recursions $n$. There is feedback from the compression stage to increase $n$ if the image complexity results in higher bandwidth requirements. The range of $n$ varies from 2–6, corresponding to low-high levels of smoothing. During this processing, the gradient image $\nabla I$ is determined via horizontal $\nabla I_H$ and vertical $\nabla I_V$ Sobel (gradient) filters.

### 3.4. Phosphenemapping

When Brindley and Lewin [18] inserted a planar array of equidistant electrodes, a pseudo-random arrangement of phosphens was returned. Similarly, when our proposed optogenetic prosthesis is utilised in

---

Table 2. Pyramidal processing function per image resolution.

| Image resolution | Total pixels | Layer 1 | Layer 2 | Layer 3 | Total processing pixels |
|------------------|--------------|---------|---------|---------|-------------------------|
| $192 \times 192$ | 36 864       | $192 \times 192$ | $96 \times 96$ | $48 \times 48$ | 11 016 |
| $128 \times 128$ | 16 384       | $128 \times 128$ | $64 \times 64$ | $32 \times 32$ | 5074 |
| $96 \times 96$  | 9216         | $96 \times 96$ | $48 \times 48$ | $24 \times 24$ | 2888 |
Figure 3. The image mapping concept. (a) An original image. (b), (c) Pyramidal scheme to reduce processing requirement. (d) Retinally processed image on pyramidal image based on the most accurate top row in table 2. (e) A discrete phosphene map with coordinates based on eccentricity (r) and radial angle (θ). (f) The original image mapped to the phosphene map. (g) The pyramidal image mapped to the phosphene map. (h) Retinally processed image mapped to the phosphene map. (i) Random map. This has been utilised to explore 2048 (i), 1024 (j) and 512 (k) stimulating phosphenes, as shown in table 1. In all cases, the distance from the centre to the periphery represents a normalised distribution within the visual field.

clinical trials, we expect to have to map each phosphene to each stimulation point. The map will then allow for determination of which points in space can be used as active phosphenes and image stimulation assigned accordingly. The information can then be rearranged to a one-dimensional stream arranged according to individual brain units. The implantable control unit will then rapidly be able to assign stimulation upon receipt.

To simulate and evaluate the outcome of our proposed phosphene mapping, we created a map of 2639 phosphenes—which equates to an optical stimulation radius of a few hundred microns per stimulator. This can be seen in figure 3(e). We then created random patterns of [2048, 1024, 512] phosphenes figures 3(i), (j) and (k) to represent potential phosphene maps.

Finally, for the scenario in which portions of V1 are inaccessible (e.g. within the calcarine sulcus), an hour-glass configuration was created for 512 phosphones as shown in figure 3(l). This is for illustration purposes, as the primary objective of this paper is to determine the main transfer of information from a headset and video processor to the internal control units and stimulators. We, therefore, chose a lower resolution of 512 phosphenes equating to 512 effective stimulators as shown in table 1. Higher-resolution implants would clearly improve the resolution within the hour-glass visual field further.

We have presented the visual field in relative rather than absolute terms. This is because it will be very dependent on the surgical approach, which is beyond the scope of this paper. A further caveat should be noted that the superior peripheral field is geographically located deeper within the longitudinal fissure between the cerebral hemispheres (i.e. more anterior portions of V1). But for reference, Brindley and Lewin [18] achieved a peripheral field of 35° in the inferior quadrant and 10° in the superior quadrant. Similarly, Dobelle and Mladejovsky [57] achieved 10° and 5° respectively.

3.5. Visual encoding

Our target for stimulation is V1 of the visual cortex. This is the region which receives information from the retinae via the LGN. There are three primary input pathways to V1: parvocellular, magnocellular and koniocellular cells in different layers of the LGN. Respectively they broadly encode, spatial-frequency, spatio-temporal information, and colours/textures. These pathways can be distinguished by their input location into V1—respectively: Layer 4C-β, Layer 4C-α, and Layer 1 [58]. For the purposes of
visual return, temporal information may be challeng- 
ing without synchronising the camera video stream to eye movement and the vestibulo-ocular reflex. As such, the target pathway would be the parvocellular pathway, which can provide spatial information.

The parvocellular output stream from the retina-
>LGNg synapses with simple cells in layer 4 of V1. These simple cells are orientation-specific, inte-
grating information from individual centre-surround outputs from the retina. They are also arranged into 
ocular dominance columns. As such, stimulus, and thus control of the firing rate of these cells must be in the form that they would transmit.

A number of studies have described the visual cortex comprising of simple and complex cells, e.g. by 
Martinez et al [29]. However, for this work, we are assuming that it would be difficult to provide targeted stimulation of each individual orientation column for each individual geographical location. As such, we constrain our stimulation to represent (±45°, ±90°, ±135°, ±180°) or alternatively we can simply provide an absolute unidirectional ON pathway derivative—
which equates an optrode stimulating multiple or all of these orientations simultaneously.

The outputs from the initial video processing are horizontal and vertical derivatives: ∇I_{H}, ∇I_{V}. 
Respectively, these can be used to determine the gradients of each of the four directions, or the abso-
lute magnitude. As these are individual scalar calcu-
lations, they can be performed after the phosphene mapping stage to reduce the number of required com-
putations. They can be calculated as follows:

\[
∇ I_{90/270°} = ± ∇ I_{H} \tag{4}
\]

\[
∇ I_{0/180°} = ± ∇ I_{V} \tag{5}
\]

\[
∇ I_{45/225°} = ± (∇ I_{H} + ∇ I_{V}) \tag{6}
\]

\[
∇ I_{135/315°} = ± (∇ I_{H} - ∇ I_{V}) \tag{7}
\]

\[
∇ I_{mag} = α \cdot max \{∇ I_{H}, ∇ I_{V}\} + β \cdot min \{∇ I_{H}, ∇ I_{V}\} \tag{8}
\]

We utilized an alpha max, beta min algorithm to improve computational efficiency. Positive/negat-
ive values are separated into each orientation vector. We then set threshold values τ_{max} and τ_{min}. Above and below these thresholds, values were set to 1 and 0, respectively. The lower threshold aids compres-
sion. The upper threshold has the effect of saturating (enhancing) important gradients.

3.6. Signal encoding and compression

As the pixel count increases, so too will data rate. 
Data for 2048 stimulators at an 8-bit range and 25 Hz 
update represents a sustained theoretical data rate of 
around 400 kbits s^{-1}. However, with error-correction 
codes and headers, the actual data rate would be closer 
to 500 kbits s^{-1}. This data rate can be challenging 
for MedRadio units. For example, Bluetooth—a com-
mon transmission protocol operating in the MedRa-
dio bands (and used in some pacemakers) is rated at 
1 Mbit s^{-1}, but typically can only achieve sustained 
data rates of around 125 kbits s^{-1}. There is, therefore, 
need for both custom high-speed data communica-
tion as well as data compression suitable for visual 
prosthesis data streams.

The most common form of video compression are 
MPEG and H.264 protocols which are analogous to 
JPEG and Motion JPEG for single images and image 
sequences. These protocols utilise the discrete cosine 
transform to convert images into frequency space and 
remove less noticeable higher frequency components. 
However, to successfully encode and decode using 
frequency methods, there needs to be a continuous 
form to the images. However, in our case, as per fig-
ures 3(i)—(l), the images have significant gaps. Fur-
thermore, pixels are not arranged in a convenient 
cartesian form, and discrete cosine transforms can be 
power-hungry.

One of the subset components of the JPEG/MPEG 
algorithm is run length encoding (RLE). This scheme 
is useful for when there is repetition, e.g. for a large 
expans of sky with similar intensity and colour. We 
have adapted this scheme to create a zero run length 
encoding (zRLE) compression method. The basis for 
this is that if we separate positive and negative gradi-
ents (ON and OFF pathways) then the image effect-
ively becomes a thresholded spatial derivative as per 
figure 3(d). As can be seen, in this case there is a sig-
nificant amount of ‘black’ (i.e. intensity = 0)—which 
is a repetition which can be removed.

The pre-encoding step is the initialisation of zero 
run length pixel stream, which is fixed to the num-
ber of active stimulators on the implant side. zRLE 
operates on the following principle: (i) the image is 
converted into a one-dimensional stream of inform-
ation with the position of each intensity value cor-
responding to a particular phosphine coordinate. As 
the pixels are a log-polar distribution rather than 
cartesian, phosphes are extracted in a spiral form 
from the centre to periphery. (ii) The encoder checks 
to see if the value is zero. If so, it checks how many 
subsequent values are also zero. In this case, it trans-
mits a byte defining the zero value, followed by a 
number < n-bits > in length corresponding to the 
repetition. (iii) The encoder then passes to the sub-
sequent non-zero value. Later the decoder places the 
non-zero values into a sequential buffer and expands 
the zero values according to the < n-bits > value. A 
diagram of this process can be seen in figure 4.
Figure 4. Zero run length encoding scheme. (a) A stream of information where n is the number of pixels in an image. After image processing, a threshold is applied, converting low importance pixels to zero. (b) Non-zero pixels are simply transmitted in binary. (c) Given the high proportion of zero pixels, when a zero pixel is defined, an additional value is provided to state how many subsequent zeros are present between $2^{n-1}$. Analysis of optimal N is provided in the supplementary data.

The compression is a function of how much ‘black’ is in the image so increasing the threshold $\tau_{\text{min}}$ in section 3.5 will set more values to zero and thus increase compression. We have explored the optimal value of N-bits with a threshold value and data size, which can be seen in the supplementary information. Generally, for small thresholds < 50 (out of 255), the optimal value is $N = 4$. For larger thresholds, the optimal value $N = 8$.

We determine the $\tau_{\text{min}}$ threshold value based on the efficacy of compression of the previous frame. We assume that the difference in the required threshold between neighbouring frames on average will not be that different.

According to the required compression rate, the threshold value $\tau_{\text{min}}$ is determined using the histogram integrating approach. Such an approach dynamically integrates the histogram of the retinal processed image until reaching the required compression rate, as such $\tau_{\text{min}}$ is the value at which the integration ends:

$$\text{Hist}(t)_{\text{Integration}} = \sum_{i=0}^{\tau_{\text{min}}} H(i)$$ (9)

The functions are performed on each pyramidal layer separately and then mapped to the phosphene map. The processing sequence for an image through the different pyramidal layers can be seen in figures 3(a)–(d). The simplification and retinal processing algorithms were designed in MATLAB and then transferred and compiled in C utilizing the OpenCV library. Using this approach, we were able to operate the processing sequence on a Raspberry Pi processor in real time.

3.7. Even power distributed pulse width stimulation

Optogenetic forms of neuromodulation involve illuminating neurons with a defined intensity for a defined time. If the stimulus time is short, then from Nikolic et al [59] and Grossman et al [60], the effective neural stimulus $S$ can be approximated to:

$$S = \int \varphi_{\mu\text{LED}} \cdot dt \approx \varphi_{\mu\text{LED}} \cdot \Delta t$$ (10)

Where $\varphi_{\mu\text{LED}}$ is the optical intensity from the LED and $\Delta t$ is the stimulus time. Note that we are ignoring the logarithmic sigmoid relationship between neural activity and which has been discussed in section 2.2. If we ignore droop effects (i.e. the LED efficiency decreases as the drive current increases) for simplicity, then we can state that the optical intensity is proportional to the LED drive current, i.e.

$$\varphi_{\mu\text{LED}} \propto I_{\mu\text{LED}} = \beta \cdot I_{\mu\text{LED}}$$

where $\beta$ is the current to light conversion factor. This conversion factor can be variable with current to include the droop factor for more accuracy.

The most convenient form of driving large numbers of LEDs is to vary the stimulus pulse width for a fixed (maximum) drive current, i.e.:

$$S(i) \approx \beta \cdot I_{\mu\text{LED}} \cdot \Delta t_{\text{PWM}}(i)$$ (11)

Where $i$ is the LED number, $\Delta t_{\text{PWM}}$ is the pulse width time and $I_{\mu\text{LED}}$ is the common drive current for all LEDs. In this case, the modulation pulse time per stimulation point can then be expressed as:

$$\Delta t_{\text{PWM}} = \frac{C_s}{(2^{15R} - 1)} \cdot t_{\text{frame}} \text{ (ms)}$$ (12)
Figure 5. Power management. (a) A stream of pixels where \( n \) is the number of pixels and \( \Delta t_{PWM} \) is optical stimulating duration. (b) The total number of activated LEDs should be less than LEDs limit and the rest of LEDs should be delayed. \( \Delta t_{Delay} \) is the time by which the LEDs is delayed. (c) The distribution of power without any power management. First \( n \) ms for stimulation and the rest of time for neural responses; Left y-axis is the sequence number of LEDs; right y-axis is the number of ON LEDs. (d) Power management simply delaying LEDs in sequence (1024 phosphenes). (e) Power management simply delaying LEDs in sequence (2048 phosphenes). (f) Power management using the advanced even power distribution method (1024 phosphenes).

Where \( C_n \) is the intensity of pixel \( n \), \( t_{frame} \) is the full stimulating time (milliseconds) and TDR is the temporal dynamic range which can vary from 4–8. i.e. we set the maximum denominator as 255 to match imaging pixel dynamic range, but it may be reduced due to the compression scheme threshold described previously. This scheme is displayed in figure 5(a). A time gap prior to the following frame can be used to allow LEDs to cool and ensure the implant surface remains below \( +2 \) °C. Also, such pulsed mode operation is optimal due to the biophysics of channelrhodopsins which have dark and light-adapted states [59, 60].

However, an issue arises in that the total current \( I_T \) at any moment \( t_x \) is the sum power of all \( \mu \)LEDs and can be expressed as:

\[
I_T(t) = \sum_{n=1}^{n=m} I_n \bigg|_{t=t_x} \quad (13)
\]

Where \( i \) is the sequence number of \( \mu \)LEDs and \( I_n \bigg|_{t=t_x} \) is the power at the moment of \( t_x \). Note that while the drive voltage can vary considerably from LED to LED, the correlation between optical output and current has very limited variation. As such, we calculate with current rather than power.

The problem with this scheme can be seen graphically in figure 5(b), i.e. as all the LEDs that will turn on do so from the start of the frame, there is an initial spike in current requirement followed by a decay. This will lead to large current fluctuations which could be undesirable form many reasons: (i) there may be a maximum drive current capability by the implant—e.g. 1 mA LED current \( \times 512 \) LEDs is a very high current for an implantable device; (ii) there may be safety concerns with large current spikes; (iii) large current spikes can cause fluctuations in heating which may surpass regulatory limits; and (iv) large current fluctuations may affect recording and diagnostic circuits.

As such, we have developed a scheme to distribute the power evenly—what we term the ‘even power distributor’. This is achieved by defining a maximum current \( I_{Max} \) and then only allowing a number of LEDs to turn on that would match that current at any given time. This can be expressed as:

\[
I_T(t) = \sum_{n=1}^{n=m} I_n \bigg|_{t=t_x} \leq I_{Max} \quad (14)
\]

This power distribution can be achieved by creating a delay time \( \Delta t_{Delay} \) for each LED. Figure 5(c) illustrates this concept. A number of LEDs are turned on until the current threshold is reached. After that, new LEDs can only turn on after other LEDs are finished. As such, the first LEDs in the sequence would turn on first eventually leading to the final LED turning on. This can be seen in figure 5(d): the number of LEDs at any moment in time is set to the maximum determined by \( I_{Max} \) and then decays once there are fewer remaining LEDs than this threshold.

However, if the integral of the maximum current within the stimulation period is less than the LED current integral, then there will be losses. This can be seen in figure 5(e), which has the same threshold...
and arrangement as figure 5(d) but now with twice as many stimulators. As can be seen, the current remains constant until the end of the frame without a decay, indicating that perhaps not all of the stimulators have been illuminated. This can be mathematically expressed as follows:

$$\text{if } \left\{ \sum_{i=0}^{t_f} \sum_{i=m}^{t_f} I_{i} \leq \left( I_{\text{Max}} \times t_f \right) \right\} \rightarrow \text{Stimulation loss}$$

(15)

Where $t_f$ is the frame time. As such, there may be situations where the final LEDs do not get illuminated, thus distorting the scene. To mitigate this, we consider the stimulation sites with the highest values to be the most important (i.e. the brightest and thus most significant spatio-temporal derivative information). As such, when the information is decoded in the implantable control unit, the sequence automatically correlates to an LED position. At this point, each non-zero value can have its location binned to set values in the number sequence. Then instead of sequentially illuminating each LED by number, each LED is illuminated in sequence according to its stimulus intensity. The outcome can be seen in figure 5(f). The distribution of pulse width stimuli now seems randomised in location. However, upon inspection, it can be seen that the longer pulses are at the beginning of the frame and the shorter pulses at the end.

4. Results

4.1. zRLE encoding results

Figure 6 shows the effect of zRLE encoding for different image sizes. (a) Shows a synthetic test image for calibration and analysis purposes. To the right are the log-polar images in different resolutions. These are [2048, 1024, 512] phosphences (P). We also show the same number of phosphences which only stimulate within the top and bottom quadrants if the calcarine sulcus proves inaccessible. We denote this HG (Hour Glass effect). These images display a zero error from a communications perspective, but simulate a pseudo-random arrangement of phosphences as would be expected from a real-world implant.

Figures 6(b) and (c) show the effect of lossy compression on image quality. i.e. as the number of stimulators increase, a higher zRLE threshold is required, which results in some low-intensity phosphences not being stimulated. Unaffected phosphences have been designated as white circles, and those affected have been designated with black filled circles. Also affecting this process is the data rate. For this work, we define two sustained (i.e. actual data excluding headers, error correction and taking into account buffering and microcontroller transmission) data rates: 125 kbits s$^{-1}$ and 250 kbits s$^{-1}$, then the maximum frame data size would be 5 kbits and 10 kbits respectively. These thresholds are shown as a red striped line. Arrows show the effect on error rate for these specific cases. Note, for the 512P case 120 kbit s$^{-1}$ is sufficient to transmit without compression.

We explored the efficacy of the zRLE in comparison to other techniques, the results of which can be seen in table 3. For all schemes, we used 3277 images of $64 \times 64$ pixels taken from the daily life routine, where scenes from inside a house, streets, and garden are encountered. The threshold of all images is set arbitrarily to 51 (20% of 255)—which is a realistic value for operational use.

RLE, Bz2 and Zlib are lossless compression techniques, while JPEG is lossy compression technique. zRLE is lossy in that the compression is determined by the threshold set for determining the zero point. Depending on the compression requirement, this may increase or decrease and can be fixed or be part of a closed-loop feedback described in figure 2. $\mu_{\text{compressed}}$ represents the mean size of the images after compression, while $\sigma_{\text{compressed}}$ is the standard
Figure 6. Zero run length encoding quality as a function of phosphene number (image size). (a) Log-polar images contain white and black phosphences to show the effects of thresholding on different resolutions (512, 1024, 2048 phosphences). Black phosphences represent those lost due to the thresholding within the zRLE compression. (b), (c) Effects are shown for a conceptual data rate of 125 kbits s\(^{-1}\) and 250 kbits s\(^{-1}\), respectively. These equate broadly to sustained data rates from bluetooth low energy 4, 5 (depending on the device). Higher data rates result in fewer losses as the total number of stimulators increase.

Figure 7. The zRLE scheme as a function of threshold value. The left y-axis (blue) refers to image frame data size in kb, while the right y-axis (purple) refers to the error rate, i.e. the percentage of pixels which deviate in value from the original processed value. From left to right is respectively results for 512, 1024 and 2048 phosphences representing three key scenarios. Thresholds are for sustained data transfer rates of 125 kbits s\(^{-1}\) and 250 kbits s\(^{-1}\), assuming 25 frames per second (red dashed lines). The blue and purple curves show the distributions of data sizes and error rates. The deeper color means more samples distributed.

Table 3. Comparison of different compression schemes.

|                      | Bz2 BWT-RLE-MTF-Huffman Coding | Zlib LZ77-Quantiﬁcation-Huffman Coding | JPEG DCT-Quant.-Huffman Coding | RLE Run Length Encoding | zRLE zero Run Length Encoding |
|----------------------|--------------------------------|----------------------------------------|-------------------------------|-------------------------|-------------------------------|
| \(\mu\) Compressed (Bytes) | 1539                           | 1677                                   | 3535                          | 2060                    | 1286                          |
| \(\sigma\) Compressed (Bytes) | 348                            | 381                                    | 733                           | 646                     | 378                           |
| Compression Time (ms) | 31.7 ± 4.7                     | 11.5 ± 2.4                             | 0.8 ± 0.1                     | 0.4 ± 0.1               | 0.3 ± 0.05                    |
deviation of the size of the images from $\mu_{\text{compressed}}$. The mean and standard deviation units are byte and the compression time in milliseconds.

The key outcome is that the zRLE compression extent and compression time are very favourable. It should be noted from a qualitative perspective that the other schemes do not produce great results with these types of images. In particular, JPEG uses a discrete cosine transform which assumes a continuous flow of pixels. However, the nature of our phosphenes images is discontinuous with gaps between pixels, which in any case are not in a convenient cartesian arrangement. As such, though simple, we find the zRLE qualitatively produces the best results.

4.2. Even power distributed PWM encoding

We developed a pulse width modulation scheme to prevent large current fluctuations as the LEDs turn on at the start of each frame. However, as per equation (15), if the multiple of frame time and the maximum current is higher than the integral of the pulse width modulations over all pixels, then there will be losses. To explore the effect of PWM management and PWM limits, we used the same 1800 frame video sequence (as shown in figure 7) to develop the results shown in figure 8.

Figures 8(a)–(c) show power distributions of [512, 1024, 2048] phosphenes using the simple power management approach. The flat line for different thresholds shows the maximum number of LEDs at any time is limited and is consistent with figure 7. Where there is a decay prior to the end of the frame, there are basically no losses. Alternatively, where the number of LEDs being utilised continues to the end of the frame, then there are losses. For example, in the case of the 512 LED limit, there is a decay function in each resolution. However, for the lower limit of 64 LEDs, no decay can be seen in the case of a 2048P image. These losses result in errors, which are presented in figure 8. (d) which shows the average error rates of 512, 124, 2048 phosphenes image as a function of logarithmic maximum power limits.

Figure 8(e) shows the LED Power/usage statistics over the course of the frame when using the advanced even power distribution scheme. Rather than show the number of ON LEDs as a function of frame time, which may show a flat power distribution, the x-axis is converted into the number of max LEDs limits. Then, the function form has become an increase, where the number of LEDs is equal to max LEDs limits, indicating information losses, prior to a flat line, where the number of LEDs is equal to its stimulus scheme at a given time and there are no losses. (f) shows the error rate of even power approach vs maximum LED utilisation for the three image resolutions.

Given that there is some lossiness in the even power distribution scheme, we have chosen an exemplar image to display the effect. Figure 9(a) shows a penguin scene with both original (top) and retinal processed (bottom) versions inserted into the simulated log-polar view; both are shown at a maximum resolution of 2639P phosphenes. (b) Provides an illustration of how they would look with simple even power distribution (top) and advanced even power distribution (bottom) for three different maximum LED thresholds [64, 128, 256] with 1024 phosphenes. Although there is no determinable difference between the two management schemes in figure 9, the types of information losses are different. Peripheral and detailed information are lost respectively by simple and advanced even power management scheme.

5. Discussion

Our approach is scalable to thousands of phosphenes, though it will take some time and significant technical effort to reach that point. In particular, effective multifork optrode probes which are sufficiently durable for long term chronic use are yet to be presented in the literature. Nevertheless, we believe our system is scalable and have implemented the external processing part on a Raspberry Pi processing unit and achieved normal video rates. Here we summarise both our novelty and challenges for the future.

We feel the system we present as a whole is novel. Furthermore, there are two key important sub-components: (i) our zRLE compression scheme; and (ii) our even power stimulation scheme.

With regards to the zRLE compression and decompression scheme; It could be argued that with a sufficiently fast wireless transceiver that such a scheme is not required. However, the decoding of the zRLE requires minimal processing. This is because our zRLE only compresses the zero elements. In contrast, variable-length encoding techniques associated with context-based detection, such as LOCO-I algorithm associated with Golomb coding [61] (which forms ITU-T recommendation T.87 standard) try to losslessly compress non-zero elements resulting in a higher compression rate. As such, a higher bandwidth transceiver would be expected to consume more energy, which is limited by thermal constraints. Furthermore, if bandwidth could be improved within the thermal boundary limits, it may be better to utilise it for closed-loop feedback control or improved temporal response and thus still use zRLE compression.

The second key innovative component is the even power distribution scheme. In earlier work [54], we showed the importance of LED efficiency and thermal constraints within stimulation. In addition, we need to consider how large swings in LED driving current could affect the system as a whole. As such, our driving scheme is broadly applicable to all forms of optogenetic prosthetics. There is a caveat to our scheme in that even power distribution within a given frame time will introduce latencies between the first LED to
Figure 8. Simple power encoding vs even power distribution. (a), (b) and (c) The power distributions using simple power managements with different LED limits in the resolutions of 512, 1024 and 2048 phosphens, respectively. The x-axis is stimulation duration (stimulation occurs in first n ms over 40 ms) and the y-axis is the number of ON LEDs. (d) The error rate of simple power management in different LED limits. The x-axis represents the log of the maximum number of LEDs allowed to be on at the same time and the y-axis represents error rates. (e) The even power management in different LED limits in the resolution of 512, 1024 and 2048 phosphens. The x-axis represents the maximum number of LEDs and the y-axis represents the number of ON LEDs. (f) The error rate in different LED limits with the resolution of 512, 1024 and 2048 phosphens. The x-axis represents the maximum number of LEDs, and the y-axis represents the error rate.

Figure 9. Power width management quality as a function of maximum LED limit thresholds. (a) The original image and its edge version with full phosphens. (b) The log-polar images with 1024 phosphens using simple (top row) or advanced (bottom row) power management schemes for different LEDs limits [64, 128 and 256].

turn on and the last. We have currently devised our scheme to operate within a 40 ms frame. However, should latencies be a problem, then that frame time can be shortened. However, this would increase the data rate and compression requirement.

Although we believe our approach to be valid, there are some caveats which require further consideration but are beyond the scope of this paper: (i) the retinal encoder; (ii) eye to camera spatial synchronisation; (iii) calcarine sulcus and hour-glass effect; and (iv) long term effects of the implant.

We have utilised a very simple spatio-temporal visual encoder to mimic the primary information that would be expected entering V1 of the visual cortex. In the future, as improvements in neural specificity and identification are made, visual encoders such as that
of Nirenberg and Chetham [35] and Jepson et al [36] could improve the ability for the brain to understand the presented information further.

Another key challenge will be the synchronisation between the visual scene presented to the cortex and eye movements. If the camera is fixed to the head direction and the eyes are moving, then it may be difficult for the visual cortex to interpret the scheme. This issue will be faced by all forms of visual brain prosthesis and some forms of retinal prosthetics. Some possible solutions include: (a) advanced eye tracking to update the visual scene that is transmitted from the camera for further processing; (b) surgically disabling the ocular muscle; or (c) inserting the camera into the eye or onto a contact lens. Each of these solutions would require significant effort and would be challenging to implement, though (a) would be the least invasive and thus preferred route. There is also a complication in that shaking of the headset, and thus the headset camera can introduce motion artefacts during walking which also need to be compensated (e.g. using inertial sensors to measure movements of the camera).

We highlight surgical-anatomical challenges of visual cortical prosthesis given that parts of the visual cortex are difficult to access (e.g. buried in the calcarine sulcus). This will make interpretation significantly more difficult. Arguably, it is the same problem for retinal prosthetics with a single implant radially positioned from the fovea. However, it is certainly a challenge for long-term consideration. We attempted to address some of these challenges in our study by using an hour-glass configuration based on [31]. It could potentially be negated by new types of implants or surgical techniques. Alternatively, it could be negated by stimulation protocols in V2 and/or V3.

Another key challenge will be the synchronisation between the visual scene presented to the cortex and the direction of the eye. If the camera is fixed, and the eye is moving, then it may be difficult for the visual cortex to interpret the scheme. This issue will be faced by all forms of visual brain prosthesis and some forms of retinal prosthetics. A simple workaround would be to ask the patient to only look forward, but that would not be ideal. Thus, possible solutions include: (a) advanced eye tracking—with matched camera movements (b) surgically disabling the ocular muscle (c) inserting the camera into the eye, or onto a contact lens. Each of these would require significant effort and would be challenging to implement, though (a) would be the least invasive and thus preferred route. There is also a furthermore minor complication in that shaking of the headset and thus the headset camera can introduce motion artefacts during walking, which also need to be compensated.

We highlight the hour-glass effect that we expect due the cortex buried in the calcarine fissure. This will make interpretation significantly more difficult. Arguably, it is the same problem for retinal prosthetics with a single implant radially positioned from the fovea. However, it is certainly a challenge for long term consideration. It could potentially be negated by new types of implants or surgical techniques. Alternatively, it could be negated by stimulus protocols in V2 and/or V3.

Finally, we acknowledge that we propose this work prior to studies that have looked at the long-term impact of the implanted high-density optical probe. Issues will include thermal management and biocompatibility, including reduced gliosis, and optimisation of the gene therapy to achieve optical targeting of specific neural sub-circuits. Thus we expect that any early clinical trials for our proposal will use more simplistic lower-resolution schemes, e.g. a single brain unit with 16 stimulation points.

6. Conclusion

In this paper, we have presented a video processing scheme for visual prosthesis, which includes video acquisition image pre-processing, mapping, compression-decompression and even power management. We believe this system as a whole has an important novelty to the progression of all forms of visual prosthetics. At a subsystem level, we have developed a zRLE compression-decompression scheme which is broadly applicable to all forms of visual prosthetics. Furthermore, we have developed an even power distribution scheme which can prevent overloading of the power management system when large numbers of LEDs need to be addressed. We have implemented the schemes to operate in real time on a Raspberry Pi single-board computer, which we will utilise for our future hardware implementation.

Acknowledgments

Dr Degenaar would like to thank the European Commission for funding the FP7 OptoNeuro project 249867, and the Wellcome Trust (102037/Z/13/Z) and the Engineering and Physical Sciences Research Council (NS/A000026/1) for supporting the related CANDO project. The authors have confirmed that any identifiable participants in this study have given their consent for publication.

ORCID iDs

Walid Al Atabany https://orcid.org/0000-0002-8141-2511
Patrick Degenaar https://orcid.org/0000-0002-5984-6698

References

[1] World Health Organisation 2014 Visual impairment and blindness Fact Sheet N° 282
[2] Edwards T L, Jolly J K, Groppe M, Barnard A R, Cottriall C L, Tolmachova T, Black G C, Webster A R, Lottery A J and
[45] Al-Atabany W I, Memon M A, Downes S M and Degenaar P A 2010 Designing and testing scene enhancement algorithms for patients with retina degenerative disorders Biomed. Eng. Online 9 27

[46] Nikolic K, Grossman N, Yan H, Drakakis E, Toumazou C and Degenaar P 2007 A non-invasive retinal prosthesis—testing the concept Conf. Proc. IEEE Eng. Med. Biol. Soc. pp 6264–7

[47] Kleinlogel S, Feldbauer K, Dempski R E, Fotis H, Wood P G, Bamann C and Bamberg E 2011 Ultra light-sensitive and fast neuronal activation with the Ca2+-permeable channelrhodopsin catch Nat. Neurosci. 14 513–U152

[48] McGovern B, Drakakis E, Neil M, O’Brian P, Corbett B, Berlinger-Palmini R and Degenaar P 2011 Individually addressable optoelectronic arrays for optogenetic neural stimulation 2011 IEEE Biomedical Circuits and Systems Conf. pp 329–32

[49] Soltan A, McGovern B, Drakakis E, Neil M, Maaskant P, Akhter M, Lee J S and Degenaar P 2017 High density, high radiance mu LED matrix for optogenetic retinal prostheses and planar neural stimulation IEEE Trans. Biomed. Circuits Syst. 11 347–59

[50] Ramezani R, Liu Y, Dekhoda F, Soltan A, Haci D, Zhao H, Firfilionis D, Constantindou T G and Degenaar P 2018 On-probe neural interface ASIC for combined electrical recording and optogenetic stimulation IEEE Trans. Biomed. Circuits Syst. (accepted) (https://doi.org/10.1109/TBCAS.2018.2818818)

[51] Barrett J M, Hilgen G and Sernagor E 2016 Dampening spontaneous activity improves the light sensitivity and spatial acuity of optogenetic retinal prosthetic responses Sci. Rep. 6 33565

[52] Holcombe A O 2009 Seeing slow and seeing fast: two limits on perception Trends Cogn. Sci. 13 216–21

[53] Castet E and Masson G S 2000 Motion perception during saccadic eye movements Nat. Neurosci. 3 177–83

[54] Dong N, Berlinger-Palmini R, Soltan A, Ponon N, O’Neil A, Travelyan A, Maaskant P, Degenaar P and Sun X H 2018 Opto-electro-thermal optimization of photonic probes for optogenetic neural stimulation J. Biophoton. 11 e201700358

[55] Al-Atabany W I, Al Yaman M and Degenaar P 2018 Extraspectral imaging for improving the perceived information presented in retinal prosthesis J. Healthc. Eng. 2018 3493826

[56] Andelson E H, Anderson C. H, Bergen J R, Burt P J and Ogden J M 1984 Pyramid methods in image processing RCA Eng. 29 33–41

[57] Dobelle W H and Mladejovsky M G 1974 Phosphenes produced by electrical-stimulation of human occipital cortex, and their application to development of a prosthesis for blind J. Physiol. London 243 553

[58] Kolb H, Fernandez E and Nelson R 2012 Webvision The Organization of the Retina and Visual System Kolb, H eds (Salt Lake City, UT: University of Utah Health Sciences Center)

[59] Nikolic K, Grossman N, Grubb M S, Burrone J, Toumazou C and Degenaar P 2011 Photocycles of Channelrhodopsin-2 Photochem. Photobiol. 85 490–11

[60] Grossman N, Nikolic K, Toumazou C and Degenaar P 2011 Modeling study of the light stimulation of a neuron cell with channelrhodopsin-2 mutants IEEE Trans. Biomed. Eng. 58 1742–51

[61] Weinberger M J, Seroussi G and Sapiro G 2000 The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS IEEE Trans. Image Process. 9 1309–24