A biologically plausible phosphene simulator for the optimization of visual cortical prostheses

Maureen van der Grinten[1]

Supervisors:
Jaap de Ruyter van Steveninck[1], Antonio Lozano[2], Umut Güçlü[1], Yagmur Güçlütürk[1]

Affiliations:
[1] *Donders Institute for Brain, Cognition and Behaviour,Radboud University, Nijmegen, The Netherlands*
[2] *Netherlands Institute for Neuroscience, Amsterdam, The Netherlands*

Corresponding author:
Maureen van der Grinten, maureen.vandergrinten@ru.nl

July 27, 2022

Blindness affects millions of people around the world, and is expected to become an increasingly prevalent condition in the years to come. For some blind individuals, cortical visual prosthetics provide a promising solution to restoring vision, by converting camera input to electrical stimulation of the cortex to bypass part of the impaired visual system. Electrical stimulation in the primary visual cortex has been found to produce dots of light in the subject's vision, called phosphenes. By evoking phosphenes in the right patterns, prosthesis wearers can be shown a representation of the outside world. As this representation has a limited resolution, visual prosthetics will need to rely on intelligent image processing algorithms that filter meaningful information from the visual surroundings. To optimise these processing strategies, non-invasive simulated prosthetic vision (SPV) can be used with sighted subjects or computational models. However, most SPV studies use highly simplified models of phosphene generation, limiting their validity for real-life applications. In this project, we developed a fast and fully differentiable phosphene simulator that transforms electrode stimulation patterns into biologically plausible representations of what the prosthesis wearer is expected to see. To achieve this, the simulator includes several computational models that take into account the visuotopic organisation of the cortex and the spread of activation in cortical tissue to determine phosphene locations and sizes. Several stimulation parameters are taken into account to model phosphene brightness and threshold values. Temporal dynamics are incorporated to allow for a realistic simulation over time. All models are parameterised and can be conveniently adapted to model empirical observations. Our results show the usability of the simulator for both computational applications as well as behavioural experiments.

Keywords: Phosphene vision, bionic vision, visual prosthesis, blindness, computational modelling, deep learning, neurotechnology, cortical stimulation

1 Introduction

As per June 2020, there are around 36 million blind people in the world. This number is expected to increase to 115 million in 2050, due to a growing global population in combination with an increase in life expectancy (Ackland, Resnikoff, & Bourne, 2017). For some blind individuals, visual prosthetics are a promising solution to restoring vision. These systems convert camera input to electrical stimulation to bypass part of the impaired visual system. Electrical stimulation in the primary visual cortex has been found to produce perceived dots of light, called phosphenes. Depending on the origin of the blindness, the electrical stimulation can take place at different points, most notably at the retina, the optic nerve, the lateral geniculate nucleus (LGN) in the thalamus or in the visual cortex.

As of 2013, retinal implants have been approved for use in humans, and in 2016 more than 350 devices had been implanted worldwide (Humayun et al., 2012; Luo & da Cruz, 2016). In this type of implant, a chip with electrodes is placed in the photoreceptor layer (subretinal), ganglion cell layer (epiretinal) or between the sclera and choroid (suprachoroidal). An advantage of such a system is that the retina is quite easily accessible compared to other parts of the visual pathway. Additionally, compared to other approaches less neural circuitry is bypassed leaving more downstream visual processing functional. However, some functionality needs to still be present in the retina for such a prosthesis to be able to work, and it is thus not an option for people with damaged retinas or problems further on in the visual pathway. Optic nerve stimulation is similarly less invasive, but it is difficult to select nerve fibres to target specific locations in the visual field due to the high amount of nerve fibres on a small surface area. The potential of LGN stimulation has been demonstrated in animal models (Panetsos, Sanchez-Jimenez, Diaz-de Cerio, Diaz-Guemes, & Sanchez, 2011; Pezaris & Reid, 2007), but so far its application has been limited by the fact that the LGN is a deep brain structure, making it harder to reach.

An alternative promising approach that has been investigated over the past decades is that of cortical implants, where electrode arrays are implanted in the (primary) visual cortex. The cortex's relatively large surface area allows for a larger number of electrodes to be implanted, increasing the number of phosphenes that can be elicited. Another advantage of cortical implants is that the system bypasses all parts of the visual pathway before the visual cortex, potentially benefiting blind people with a wider range of causes of blindness. Moreover, it seems that there is a longer window after disease onset for therapeutical intervention compared to other implant locations (Farnum & Pelled, 2020). Although the folded nature of the cortex poses a surgical challenge, the central receptive fields, most likely to be the prime target of stimulation, are located near the surface of the visual cortex allowing for easier access.

The cortex can be stimulated with different electrode types: through subdural surface electrodes located on top of the cortex, or intracortically through penetrating electrodes. Surface electrodes are easier to implant

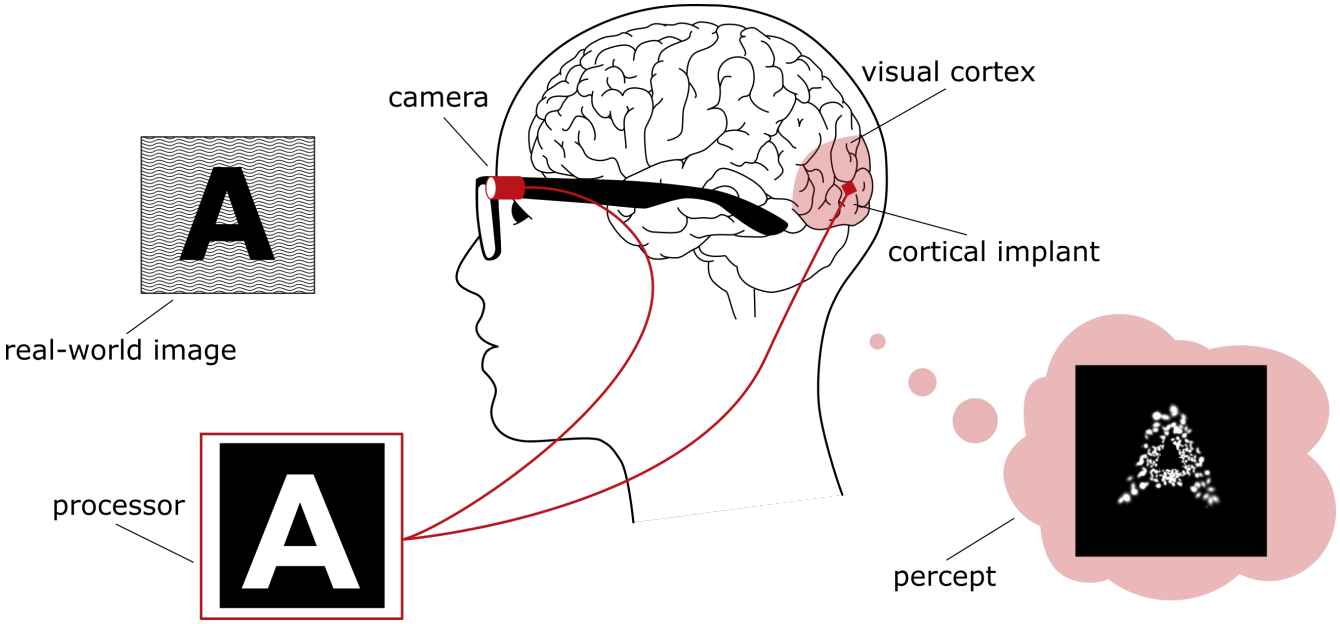


Figure 1: *Schematic overview of a cortical neuroprosthesis. A camera captures the environment, which is processed to create electrical stimulation in a cortical implant. This stimulation results in phosphene vision.*

and have for instance already been approved for medical implantation in epilepsy patients. However, surface electrodes require larger amounts of current to reach stimulation thresholds. They also lead to the perception of larger phosphenes, decreasing the potential resolution of the prosthesis. Intracortical depth electrodes, on the other hand, require lower amounts of current to elicit phosphenes, thus leading to a smaller spread of current and smaller phosphenes. This higher precision in turn allows for a higher potential resolution.

For the visual prosthesis to work, camera images need to be translated into electrical stimulation patterns to elicit a limited number of phosphenes. Current approaches yield resolutions of approximately 100-1000 phosphenes (Fernández et al., 2021; Chen, Wang, Fernandez, & Roelfsema, 2020). To be able to transfer useful information to the user at such a low resolution, image processing or feature extraction methods need to be developed that can extract meaningful information from the camera input. To optimise and test such processing strategies in a non-invasive manner, without the need for blind volunteers, simulated prosthetic vision (SPV) can be used with sighted subjects or computational models. By using simulators of phosphene vision, this line of research can accelerate the developments in the tailoring of image processing techniques to specific tasks or constraints. Phosphene simulators used in such studies transform electrode stimulation patterns into a simulated visual representation of what the prosthesis wearer is expected to see. Most SPV studies make use of highly simplified simulations of phosphene perception, often assuming that phosphenes are all equally sized, appear in a grid-like distribution, and not considering temporal dynamics (e.g., Dagnelie et al., 2007; Vergnieux, Macé, & Jouffrais, 2017; de Ruyter van Steveninck, Güçlü, van Wezel, & van Gerven, 2022; Parikh, Itti, Humayun, & Weiland, 2013). While such simplified approaches work for fast prototyping and ease the computational cost of running a simulator, the validity of these results regarding the clinical benefits of visual prosthetics is limited.

In the current project, we developed a fast and fully differentiable phosphene simulator that transforms electrode stimulation patterns into biologically plausible representations of phosphene vision. To achieve this, the simulator includes several computational models that take into account the visuotopic organisation and cortical magnification of the cortex. Our simulation models the expected spread of activation in the cortical tissue to determine phosphene sizes, and converts several stimulation parameters into a measure of effective accumulated charge to determine phosphene brightness and threshold values. Temporal dynamics are incorporated to allow for a realistic simulation over time. All models are parameterised and can be conveniently adapted to model empirical observations.

The biological plausibility of the simulator is validated on experimental data from previous clinical studies, further demonstrating the model's ability to produce expected output based on physiologically and phenomenologically grounded parameters. To showcase the possible applications of the differentiable framework, a previously proposed end-to-end model (de Ruyter van Steveninck, Güçlü, et al., 2022) is applied, which uses a fully differentiable pipeline to train an encoder model to predict stimulation parameters and a decoder model to reconstruct the original input. Using this model, several computational experiments are conducted. Firstly, the end-to-end model is trained on static images, using our simulator between the encoder and decoder models. Secondly, the trained pipeline is limited to predict stimulation protocols in realistic and safe parameter ranges. These results demonstrate the possible application of this simulator to explore realistic stimulation parameters spaces without the need for blind volunteers undergoing invasive surgical procedures. Finally, the architecture is extended to the temporal domain, demonstrating the possibilities for the optimisation of stimulation patterns over time.

Compared to the aforementioned SPV work, our experiments explore a more biologically grounded simulation of phosphene locations, sizes and dynamics. Furthermore, our simulator includes a biophysical model for processing different stimulation parameters such as stimulation current, pulse width, and frequency. Compared to other studies, where more abstract or qualitative descriptions of the required stimulation ('on' 'off') are used, this approach allows us to predict the effect of continuous (rather than binary) stimulation in realistic ranges.

1.1 Background

1.1.1 Simulated Phosphene Vision

Simulated Phosphene Vision (SPV) is often used to develop, optimise, and test processing strategies that convert raw camera input into a representation that can be used for stimulation. Studies investigating such processing strategies typically focus on one task or setting, such as navigation (Cha, Horch, & Normann, 1992; Dagnelie et al., 2007; de Ruyter van Steveninck, van Gestel, et al., 2022; Han, Srivastava, Xu, Klein, & Beyeler, 2021; Vergnieux, Macé, & Jouffrais, 2014; Vergnieux et al., 2017), hand-eye coordination (Dagnelie, Walter, & Yang, 2006; N. R. Srivastava, Troyk, & Dagnelie, 2009), reading (Cha, Horch, Normann, & Boman, 1992; Sommerhalder et al., 2003, 2004) or face recognition (Bollen, Güçlü, van Wezel, van Gerven, & Güçlütürk, 2019; Thompson, Barnett, Humayun, & Dagnelie, 2003). As research on visual prostheses has largely been focused on retinal

prostheses over the past decades, it is unsurprising that many of these studies focus on the simulation of retinal prosthetic vision.

SPV studies have yielded valuable information on strategies to condense or filter information in a manner that is interpretable to a visual prosthesis wearer. For instance, Vergnieux et al. (2017) found that limiting the information content in SPV to 6 or 9 meters vision distance improved performance in navigation tasks. Several studies have looked into the effect of the number of phosphenes, spacing between phosphenes and the visual angle over which the phosphenes are spread (e.g., de Ruyter van Steveninck, van Gestel, et al., 2022; Thorn, Migliorini, & Ghezzi, 2020; Sanchez-Garcia, Morollon-Ruiz, Martinez-Cantin, Guerrero, & Fernandez-Jover, 2022; N. R. Srivastava et al., 2009; Dagnelie et al., 2007; Parikh et al., 2013). The results of these studies vary widely, which could be explained by the difference in task complexity, but perhaps even more importantly, by the differences in how phosphene vision is simulated. Most of the aforementioned studies use highly simplified phosphene simulations, where equally-sized phosphenes are uniformly distributed over the visual field. In some cases, phosphene vision is simulated as a pixelised representation of a black-and-white version of the input, with square phosphenes in grayscale (e.g., Sommerhalder et al., 2003). In other cases, SPV was attained by using a perforated mask over a head-mounted television monitor (e.g., Cha, Horch, & Normann, 1992). The question is, therefore, whether the results of these studies generalise to the experiences of real-life prosthesis wearers. It has been stated before that more realistic simulations can close, or at the very least narrow this gap with reality (Han et al., 2021; Dagnelie, 2008).

An example of an existing realistic and biologically plausible simulator for retinal prostheses is Pulse2Percept (Beyeler, Boynton, Fine, & Rokem, 2017), which takes into account the axonal spread of activation along ganglion cells and temporal nonlinearities to construct plausible simulations of stimulation patterns. In the field of cortical prostheses, several studies have made steps towards a more realistic simulator, particularly by taking into account cortical magnification in the distribution and sizes of phosphenes (N. R. Srivastava et al., 2009) or using visuotopic maps to determine phosphene appearance (Fehervari, Matsuoka, Okuno, & Yagi, 2010; Li, 2013; N. R. Srivastava et al., 2009). However, there are still some improvements to be made, especially with regard to the effect of changes in stimulation parameters and the temporal dynamics of phosphene perception.

With a more realistic simulator, clinical experiments with sighted participants will mirror the real-life equivalent more closely. For instance, simpler phosphene simulators modelling phosphenes as uniform grids of equally-sized dots will not provide us with information on how prosthesis wearers will manage with the relatively lower resolution in the periphery compared to the foveal visual field, as caused by the organisation of the visual cortex. While more simplistic phosphene simulators may be able to provide us with informative encoding schemes, if the corresponding electrical stimulation parameters are not considered, they do not provide us with information on how to achieve these phosphene encodings. Realistic SPV, on the other hand, can aid technological developments by allowing engineers and clinicians to test the effects of changes to stimulation protocols, and subsequently select stimulation parameters that yield the desired phosphene percepts without the need for extensive testing in blind volunteers with neural implants. Finally, realistic simulators can be used as support in the rehabilitation process, to assist clinicians in identifying potential problematic situations and adapt preprocessing or stimulation protocols accordingly (Dagnelie, 2008).

1.1.2 Application of Machine Learning

The most straightforward approach to the problem of converting highly detailed camera input into low-resolution phosphene representations or stimulation patterns is to only use the outlines of the objects in the field of view, for example by using edge detection algorithms which simply look at high-contrast transitions in the input. While this is an efficient way to discard most non-essential details from the visual input, it is no guarantee that the information that remains is useful for the task at hand. To process the camera input in a smarter way, possibly tailored to meet task-specific demands, machine learning methods can be used. Examples of existing machine learning-based approaches are using emotion recognition for specialised processing of faces (Bollen et al., 2019), deep learning-based visual saliency, depth and object segmentation extraction for scene simplification (Han et al., 2021), semantic and structural segmentation for object recognition (Sanchez-Garcia, Martinez-Cantin, & Guerrero, 2020), deep learning-based surface boundary detection for mobility tasks (de Ruyter van Steveninck, van Gestel, et al., 2022) or more generally applicable end-to-end optimisation architectures (de Ruyter van Steveninck, Güçlü, et al., 2022). Other approaches leverage deep learning to optimise processing strategies that mirror biological vision (Lozano et al., 2020).

The advantage of machine learning-based methods is clear: more intelligent and flexible extraction of useful information in camera input leads to less noise or non-essential information in the low-resolution phosphene representation, allowing for more successful completion of tasks. However, as these processing algorithms will need to be able to run in real time on wearable devices using low amounts of power, it is important to consider

the computational demands of deep learning-based approaches.

The combination of machine learning frameworks for image-to-stimulation processing and realistic phosphene simulation opens new doors for optimisation methods. Although technological developments advance the state-of-the-art hardware capabilities rapidly, cortical prosthesis devices will be operating under constraints, due to both hardware limitations as well as safety limits regarding neurostimulation. Deep learning methods trained in tandem with a biologically plausible phosphene simulator can be leveraged to produce constrained optimal stimulation paradigms that take into account these limitations, allowing for safe and viable stimulation protocols to be developed.

2 Methods

The biologically plausible simulator is implemented in Python, using the PyTorch differentiable programming framework (Paszke et al., 2019). Given a specific parameter configuration, the simulator is initialised with a list of phosphene locations expressed in polar coordinates. These phosphene locations can either be initialised by defining electrode locations on a visuotopic map of the primary visual cortex, or pseudo-randomly and agnostic to the electrode locations using a probabilistic model. During the simulation, at each time step the simulator takes a set of stimulation parameters (stimulation amplitude, pulse width and frequency) and through several computational models and non-linear equations determines the appearance characteristics of each phosphene. Aspects that are computed are the phosphene size, brightness, thresholds and temporal dynamics. Each of the processing steps is founded on results in the literature from clinical and computational studies. In the following sections, the relevant findings from the literature will be discussed for each of the models in this pipeline, followed by a description of the implementation that was based on these findings.

2.1 Phosphene Placement

The primary visual cortex is organised as a map of the visual field, where neurons that respond to similar regions in the visual field will be located together. Additionally, the centre of vision is represented by a relatively larger number of neurons, while less cortical area is dedicated to the processing of peripheral vision, a phenomenon called cortical magnification (Wandell, Dumoulin, & Brewer, 2007).

To determine the locations of phosphenes in the visual field, visuotopic maps of V1 can be used. These maps have been constructed to provide a translation from visual field coordinates to locations in the (flattened) visual cortex. By using the inverse of these models, we can relate electrode locations back to the expected locations of phosphenes.

The arguably simplest visuotopic map is the monopole model (Schwartz, 1983), in which a log-polar form is used to model the inverse of the approximately linear nature of the relation between the cortical magnification factor and eccentricity:

$$w = k \log(z + a) \tag{1}$$

where w is the location in (one hemisphere of) V1 and z is the corresponding location in the corresponding half of the visual field, both in complex coordinates. The parameter a shifts the singularity of the complex logarithm along the horizontal axis, and k is a scaling factor that scales the mapping to realistic proportions in millimetres cortical distance.

The cortical magnification factor M can be found as the derivative of this equation evaluated along the horizontal meridian:

$$M = \frac{k}{E + a} \tag{2}$$

Where E is the eccentricity of the coordinate in the visual field in degrees of visual angle. M is given in millimetres of cortical surface per degree of visual angle. The inverse of the monopole mapping, translating map coordinates back to the visual field, is given by

$$z = \exp(w/k) - a \tag{3}$$

The monopole model reflects the relative magnification in the central part of the visual field. Moreover, the shape roughly matches that of a flattened V1 (Horton & Hoyt, 1991). However, while the cortical magnification factor is approximately linear in the central visual field, it has been found to be supra-linear in the periphery (Polimeni, Balasubramanian, & Schwartz, 2006). To account for this supra-linearity, the dipole model adds a second singularity at location b in the periphery:

$$w = k\,(log(z + a) - log(z + b)) \tag{4}$$

The inverse of which is given by

$$z = \frac{ab\,(\exp(w/k) - 1)}{b - a\exp(w/k)} \tag{5}$$

The dipole mapping yields a magnification factor of

$$M = \frac{k\,(b - a)}{(E + a)\,(E + b)} \tag{6}$$

which reflects the aforementioned supra-linearity in the periphery of the visual field.

When examining the organisation of V1 in combination with the areas V2 and V3, a topographic shear can be observed which is not present in the dipole mapping. To account for this, the wedge-dipole model (Polimeni et al., 2006) introduces a shearing operation in addition to the dipole model. While the full wedge-dipole model covers the full V1, V2 and V3 areas, we only regard the model for the V1 area here, as this would be the targeted site for electrode implantation.

For the transform of the wedge-dipole model, we represent the coordinates of locations in the visual field in complex polar coordinates $z = re^{i\theta}$, where r is the eccentricity of the point and $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ is the angle or azimuth. The wedge map of the visual field is then given by

$$\Lambda(re^{i\theta}) = re^{i\alpha\theta} \tag{7}$$

Where α is the parameter controlling the amount of shear. This wedge map in complex coordinates is mapped to cortex coordinates using the dipole mapping (Equation 4). Note that for V1, when a shear parameter α of 1 is used, this model is identical to the dipole model. As the shearing operation does not impact the horizontal meridian of the V1 model, the cortical magnification factor is also the same as for the dipole model.

2.1.1 Implementation

All three of the above discussed visuotopic models are implemented in the simulator and can be selected as needed. The parameters of the models are configurable, with the default values set to $a = 0.75, k = 17.3, b = 180, \alpha = 0.95$ based on a fit by Polimeni et al. (2006) on data of the human V1 area from Horton and Hoyt (1991). Electrode locations can be specified in terms of millimetres of cortical distances (see Figure 2). To simulate biological irregularities and inter-person cortical variabilities, spatial noise and dropout are applied to the virtual electrodes. The cortical magnification for each phosphene is determined based on the phosphene's eccentricity and the selected visuotopic mapping.

2.2 Stimulation Parameters

Intracortical stimulation is applied through biphasic pulse trains, defined by a stimulation current amplitude (in μA), pulse width (μs), frequency (Hz) and train duration (ms). Figure 3 shows a graphical representation of such a pulse train and its parameters. In a cathodic-first pulse train, the first pulse of a pulse-pair is negative, while it is positive in an anodic-first paradigm. Schmidt et al. (1996) found that cathodic-first pulse trains lead to lower current thresholds for phosphene elicitation (Schmidt et al., 1996), although Fernández et al. (2021) found no such effect.

Typically, variations in the stimulation current amplitude are used to elicit and in some cases control the brightness level of phosphenes, with higher stimulation amplitudes leading to higher detection probabilities and brighter phosphenes (Schmidt et al., 1996; Fernández et al., 2021). Several studies report that changes in pulse width, frequency and train duration affect phosphene appearance as well, particularly threshold values and brightness (Brindley & Lewin, 1968; Fernández et al., 2021; Niketeghad et al., 2019; Schmidt et al., 1996),

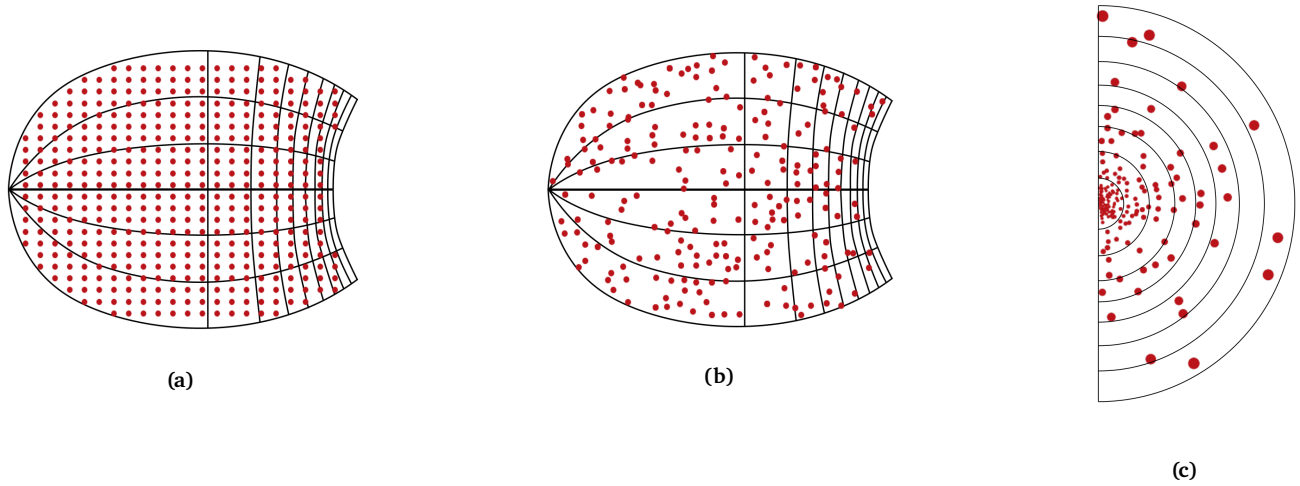


Figure 2: *Mapping from electrode locations to phosphenes.* ***(a)*** *A grid of electrodes is initialised on a visuotopic dipole map.* ***(b)*** *spatial noise and dropout are applied.* ***(c)*** *using the inverse visuotopic map, the electrode coordinates are translated to visual field coordinates. The cortical magnification is used to determine relative phosphene sizes.*

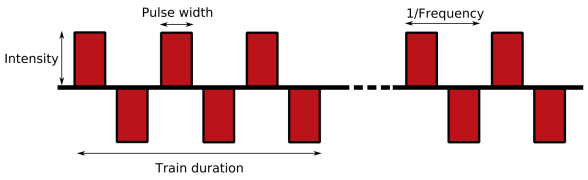


Figure 3: *Typical biphasic, anodic-first stimulation train.*

with longer pulse widths and higher frequencies leading to lower detection thresholds and brighter phosphenes. Taken together, these results imply that the detection probability and phosphene brightness are a function of deposited charge, rather than just injected current.

2.2.1 Implementation

At each time step, the simulator takes a stimulation amplitude, pulse width and frequency value for each electrode-phosphene pair. In absence of the latter two parameters, default values are provided. To incorporate the effects of changes in pulse width and frequency, a *charge per second* is calculated as the product of stimulation intensity, pulse width and frequency, in units of C/s. This scalar value summarises the rate at which charge is deposited in the cortical tissue for each phosphene.

As the train duration is in the same order of milliseconds as the frame rate of the simulator, this stimulation parameter is not taken into account. Rather, the number of consecutive frames in which a phosphene is stimulated can be regarded as a measure of the pulse duration of the stimulation train.

2.3 Phosphene Thresholds

A minimum amount of current is required to elicit phosphenes, although this minimum current depends on the specific parameters of the stimulation train. In clinical studies, these thresholds are quantified by fitting psychometric curves to experimental data, where the threshold is defined to be the amount of injected current (or charge) for which the subject perceives a phosphene 50% of the time.

Several stimulation parameters have been found to affect these threshold curves, including the pulse width and frequency of the stimulation train (Niketeghad et al., 2019; Fernández et al., 2021; Schmidt et al., 1996). Other aspects influence these threshold values further, such as the location and impedance of electrodes (Foroushani, Pack, & Sawan, 2018).

Larger pulse widths and higher frequencies consistently result in lower current thresholds, which is to be expected due to the higher charge deposited in the cortical tissue. However, when expressing the thresholds in terms of the total *charge* required to reach the threshold, the threshold increases for higher pulse widths and frequencies (Niketeghad et al., 2019; Fernández et al., 2021), implying that the efficiency of the stimulation decreases when increasing these parameters. While for lower pulse widths and frequencies a higher current amplitude is required to elicit a phosphene, the total charge that is needed will be lower. Interestingly, the threshold versus pulse width and threshold versus frequency curves established by Niketeghad et al. (2019) resemble strength-duration curves, describing the minimal current or charge injection required to excite neural tissue (Geddes, 2004). In terms of current, this curve can be described as

$$I = b(1 + \frac{c}{d}) \tag{8}$$

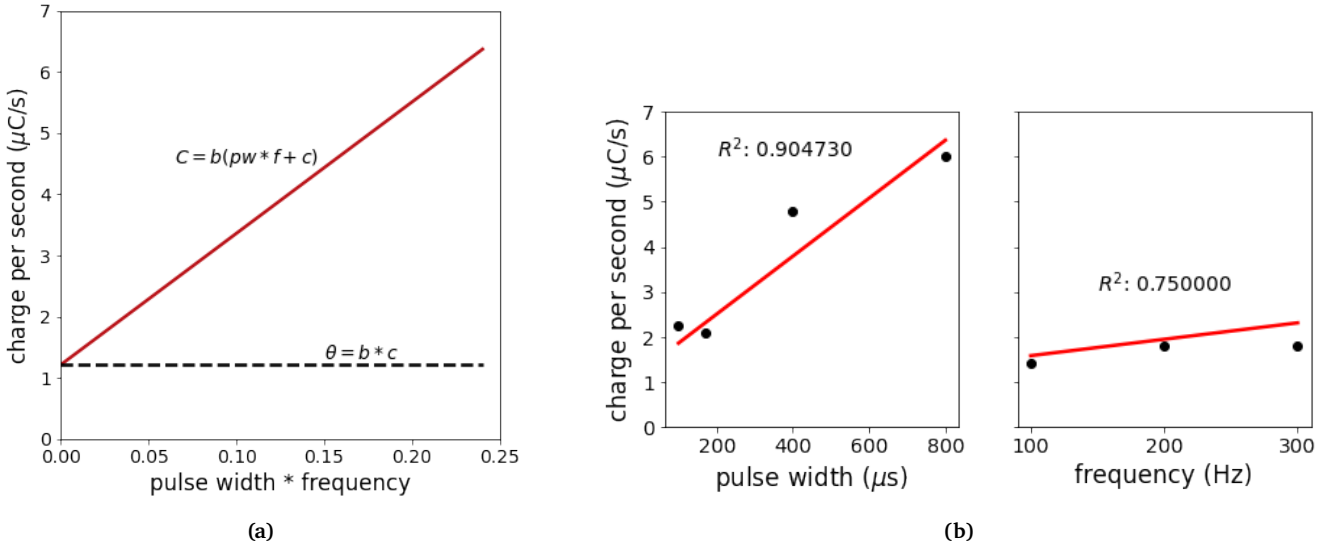
where I is the minimal required current to elicit a neuronal response and d is the duration of the stimulus. b is the rheobase value, the minimal current intensity at which a response can be elicited at infinite stimulation durations, and c is the chronaxie, the duration at which the minimal required intensity is twice the rheobase value.

This decrease in current thresholds but increase in total required charge has not only been observed in stimulation in the primary visual cortex, but in other cortex areas as well, such as the somatosensory cortex (Kim et al., 2015) and the auditory nerve (Bruce, White, et al., 1999). Computational models have been proposed that capture this behaviour, using leaky-integrator dynamics (Fridman, Blair, Blaisdell, & Judy, 2010) or neuronal firing models (Kim, Callier, & Bensmaia, 2016; Bruce, Irlicht, et al., 1999).

Due to safety considerations, electrical stimulation of cortical stimulation is typically optimised to use minimum charge and current injections while still yielding reliable results (Shannon, 1992). In light of the dynamics described above, it will be beneficial to have an accurate model of phosphene thresholds to optimise both the required current and required charge by varying the pulse width and frequency parameters.

2.3.1 Implementation

To model the variations in thresholds as observed during clinical studies, a charge per second equivalent to equation 8 is used. As both pulse width and frequency effects need to be incorporated, a multivariate version of this



(a) **(b)**

Figure 4: ***(a)*** *Charge-duration curve and the minimal threshold for maximum efficiency. The difference between the two determines the loss in efficiency that occurs for higher pulse widths and frequency.* ***(b)*** *Fit of Equation 9 on 50% detection thresholds established in Fernández et al. (2021).* ***left*** *shows the fit for different pulse widths,* ***right*** *for different frequencies.*

curve is needed. Several possible parameterisations of this equation are possible, but for the sake of simplicity the curve is defined as equation 8 where the duration d is replaced by the product of the pulse width and frequency (yielding the fraction of time during which stimulation took place). The charge per second equivalent is then found by multiplying the current with the pulse width and frequency, yielding

$$Q(\text{pw}, \text{f}) = b(\text{pw} * \text{f} + c) \tag{9}$$

where Q is the threshold in charge per second (C/s). This is a linearly increasing equation, where the intersection with the y-axis provides us with the theoretically most efficient minimal threshold θ for infinitely small pulse width and frequency. The difference between the charge per second threshold, given by equation 9, and θ gives us an approximation of the inefficiency that occurs for higher pulse widths and frequencies, in units of coulomb per second. This inefficiency could be seen as a rate of loss of charge in the underlying neural population, for instance due to leaky-integrator dynamics.

For each phosphene, at each times step, given a *deposited* charge per second input, we can use this measure of inefficiency to calculate the *effective* charge per second. This is done by subtracting the inefficiency from the total deposited input charge per second as determined in section 2.2. As this effective charge per second codes for the amount of charge per second that the targeted neural population actually receives and responds to, it is used in subsequent steps to determine thresholds and, taking into account temporal effects, scale the phosphene's brightness.

To implement a probabilistic approach to the phosphene thresholds, a logistic function is fit to the theoretical minimal threshold θ. The probability of eliciting a phosphene is thus given by

$$\rho(\text{phosphene}) = S(s * EC - \theta), \quad S(x) = \frac{1}{1 + e^{-x}}$$

where s is the slope of the sigmoid curve and EC is the effective charge per second.

The values for b and c in Equation 9 are determined by fitting this curve on the 50% phosphene perception probability values established by Fernández et al. (2021, Figure 2a,b) for a range of pulse width and frequency values (see Figure 4b). This yields a rheobase value of $b = 23.9\mu$C/s and a chronaxie value of $c = 44.5 \cdot 10^{-3}$.

The variability in threshold values between electrodes is modelled by initialising each phosphene in the simulator with its own threshold. The distribution of these thresholds follows a normal distribution with a mean value of θ and a standard deviation of 0.5μC/s, matching values reported by Fernández et al. (2021, p. 4). The slopes of the psychometric curves are initialised following a uniform distribution ($0.1 \cdot 10^7 < s < 0.5 \cdot 10^7$).

2.4 Temporal Dynamics

Behavioural data indicate that there are several temporal processes underlying the dynamics of phosphene perception.

Schmidt et al. (1996) conducted extensive experiments examining phosphene durations and accomoda-

tion in response to stimulation over time. For individual stimulation trains, they found that phosphenes had a slightly delayed onset. For short stimulation trains, the estimated phosphenes durations were slightly longer than the train length (TL), implying a delayed offset as well (estimated phosphene durations were 259, 276 and 481 ms for a train duration of 250ms). For longer TL, however, phosphenes disappeared before the stimulation had ended (i.e. phosphene duration never exceeded 930 ms for TLs of 1000 or 1500 ms). With interrupted stimulation (TL of 125 ms, inter-train-interval of 25 ms, for a total stimulation time of 1925 ms), the phosphene duration was extended, with the subject reporting seeing the phosphene for 2200 ms.

At a larger time scale, accomodation in response to repeated stimulation was examined. Over the course of 200 seconds, 50 stimulations were applied at 4s intervals. Over these 200 seconds, brightness gradually decreased to around 20% of the initial brightness of the phosphene. Recovery was examined by stimulating once every 4 minutes after the first 50 stimulations and recording brightness. Over these 16 minutes, no recovery was noted.

It should be noted that there are some disagreements in the literature regarding the temporal dynamics displayed by phosphenes. Bak et al. (1990) noted that in their experiments there was a delayed on- and offset for surface electrodes, but did not find this effect for intracortical electrodes, noting a more 'crisp' percept. Bartlett, Robert W. Doty, Lee, Negrão, and William H. Overman (1977) found habituation effects in experiments with macaques, but only after continual stimulation for 1-8 hours. Other studies found phosphenes fading after 10-15 seconds of continuous stimulation (Dobelle & Mladejovsky, 1974) or no habituation effects at all for repeated stimulation (Fernández et al., 2021).

2.4.1 Implementation

Seeing as the literature does not quite agree on the nature of accommodation and other temporal dynamics in phosphene perception, we opted to use a flexible model that can be tailored according to need. As Schmidt et al. (1996) provided the most extensive data with regards to temporal effects, their psychophysical data were used to fit the temporal models. Three temporal effects are modelled in the simulator: the delayed on- and offset, accommodation and recovery in response to continuous stimulation and accommodation and recovery in response to repeated stimulation over longer periods of time.

The delayed on- and offset can be achieved through a feedback connection in the activation (defined as the accumulated effective charge per second, see section 2.3.1) of the phosphene, mirroring leaky-integrator dynamics:

$$\text{activation}_t = \text{stimulation}_t + a * \text{activation}_{t-1} \quad (10)$$

where t is the step in the simulator and $a = 32.1/s$ is a parameter determining the strength of the feedback connection.

Both accommodation processes can be modelled by a simple habituation model (Staddon, 1993). This model uses a scalar memory *trace* that functions as an integrator over the phosphene's activation history, increasing when there is a stimulation, and decaying when there is not:

$$\text{trace}_t = b * \text{trace}_{t-1} + c * \text{stimulation}_t \quad (11)$$

where t again is the step in the simulator and b and c determine respectively the decay and increase of the trace, and with that the strength of accommodation of the phosphene. This integration operation allows the simulator to incorporate temporal effects over longer time scales without needing to save the phosphene's entire activation history. The response of the phosphene at each time step is then determined by subtracting the trace from the input activation

$$\text{activation}_t = \text{stimulation}_t - \text{trace}_t \quad (12)$$

As mentioned before, two separate processes of accommodation at different time scales have been observed. This can simply be achieved by maintaining two traces, one with faster dynamics for accommodation in response to continuous stimulation ($b_{fast} = 7.81/s, c_{fast} = 1.4/s$), and one with slower dynamics which increases for repeated stimulation over longer time spans ($b_{slow} = 0.001/s, c_{fast} = 0.14/s$). The resulting trace values can be added to form a single trace value to be used in equation 12. Figure 5 gives a visual demonstration of how this double trace process leads to accommodation in phosphene brightness.

It should be noted that the parameters a, b and c in equations 10 and 11 depend on the size of the time step. To allow different frame rates in the simulator, the model takes these parameters in terms of rates of decay or increase, and uses the frame time to establish absolute increase and decay values for the equations.

2.5 Phosphene Size

Establishing the sizes of phosphenes in blind participants in an objective and quantitative manner has, for obvious reasons, turned out not to be a straightforward endeavour. In the literature, one often comes across descriptions of sizes is terms like 'pinpoints of light' (Fernández et al., 2021), 'a nickel at arm's length' (Schmidt et al., 1996) or 'a star in the sky' (Dobelle & Mladejovsky, 1974). It has been found that higher current amplitudes elicit

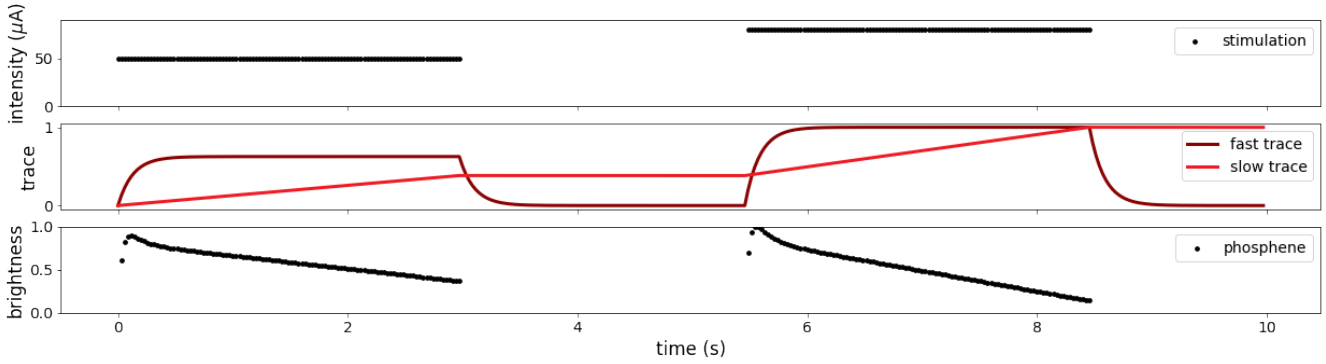


Figure 5: *The temporal model of accommodation using traces.* ***Top*** *stimulation input: continuous stimulation over 3 seconds, at intermediate and subsequently high stimulation strength.* ***Middle*** *the two traces, showing increase when there is incoming stimulation and decay when there is not.* ***Bottom*** *the resulting phosphene brightness, displaying accommodation.*

larger phosphenes (Fernández et al., 2021; Bosking et al., 2017), although a saturation in phosphene size for higher current amplitudes has been reported for surface electrodes (Bosking et al., 2017; Winawer & Parvizi, 2016). Quite consistently, surface electrodes seem to evoke larger phosphenes than intracortical electrodes, as would be expected due to the larger current amplitudes required to elicit phosphenes, and consequently the larger spread of activation in the cortex.

Several attempts have been made to quantitatively describe phosphene sizes as a function of current. Notably, Tehovnik and Slocum (2007) pioneered a method in sighted non-human primates where a stimulation-induced delay in saccades is used to map current spread for different stimulation strengths. Their results were fit with the following equation:

$$r = (I/K)^{1/2} \tag{13}$$

where r is the radius of the current spread (in mm), I is the stimulation current (in μA) and K is a spread parameter (μA/mm^2). By multiplying the cortical spread with the inverse of the cortical magnification factor M, coding the amount of tissue that represents one degree of vision, we can get an estimate of the perceived phosphene size as a function of stimulation amplitudes:

$$PS = D * \frac{1}{M} \tag{14}$$

where D is the diameter of cortical spread, and the phosphene size PS is given in degrees.

In a similar vein, Bosking et al. (2017) investigated the effect of input current on the perceived phosphene size using surface electrodes in sighted participants. Phosphene sizes were established by having the participants draw ellipses at a fixed distance from their eyes. They found that using a higher stimulation current amplitude leads to larger phosphenes, confirming earlier results, but also that this relationship saturates for higher phosphene values. By fitting a sigmoidal curve to these experimentally determined phosphene sizes, they were able to establish a quantitative relationship between this stimulation current and the spread of activation in the cortex:

$$AC = MD / \left(1 + e^{(-s(I - I_{50}))}\right) \tag{15}$$

Where AC is the diameter of activated cortex, MD is the diameter of activated cortex at which the current spread will saturate, s is the maximum increase or slope of activated cortex for increase in stimulation current, I is the stimulation strength in μA and I_{50} is the current strength at which half of the maximum current spread is reached.

2.5.1 Implementation

To determine the size of phosphenes in the simulator, the expected spread of current in the cortex is multiplied with the inverse cortical magnification factor as derived from the visuotopic map in section 2.1 (Bosking et al., 2017; Winawer & Parvizi, 2016). To obtain the expected spread of activation, either equation 13 or 15 can be used. For equation 13, the parameter K was set to 675 $\mu A/mm^2$ following Tehovnik, Tolias, Sultan, Slocum, and Logothetis (2006). In the case of equation 15, where the parameters in the literature were fit on data from surface electrodes, parameters were chosen that yielded a saturation curve over the $0 - 100\mu A$ range that is typically used for intracortical stimulation ($I_{50} = 40\mu$A, $s = 0.08$mm/μA). An approximate maximum spread diameter of $0.7mm$ was determined based on reports that phosphenes closer together yield one phosphene (Schmidt et al., 1996).

As equation 13 was originally designed for intracortical electrodes, this model is used in all results presented

in this report. However, the simulator does include the option to use the saturation curve as found by Bosking et al. (2017).

There is little data on the effect of other stimulation parameters on phosphene size. Hence, only the stimulation intensity is used to establish phosphene sizes.

2.6 Phosphene Appearance

Phosphenes evoked through intracortical stimulation have generally been described as circular white or grayish dots of light (Schmidt et al., 1996; Bak et al., 1990; Fernández et al., 2021), although more elongated shapes have been reported as well (Bosking et al., 2017; Winawer & Parvizi, 2016; Bak et al., 1990). In some cases, phosphenes were reported to be coloured (Bak et al., 1990; Schmidt et al., 1996).

As mentioned before, the brightness of phosphenes can be modulated by varying stimulation parameters (Schmidt et al., 1996; Fernández et al., 2021). Fernández et al. (2021) established curves relating stimulation amplitude to phosphene brightness, showing a saturation in brightness for higher stimulation currents.

2.6.1 Implementation

The simulated phosphenes are generated as two-dimensional Gaussian 'blobs' to produce a circular percept which fades out towards the border. The size of each phosphene is determined according to either equation 13 or 15, which yield a radius in degrees of visual angle. To translate this radius to the Gaussian percept, the parameter σ is set to half the radius, resulting in 95% of the Gaussian's activation being inside the phosphene's radius.

To introduce some noise in the perceived shapes of phosphenes, the simulator includes the option to pass a Gabor filter with random orientation over each of these Gaussian blobs, resulting in slightly elongated phosphenes. Figure 6 shows a Gaussian and a Gabor-filtered phosphene.

The height of the Gaussian determines its brightness, and is given by the accumulated activation in units of effective charge per second as described in sections 2.2, 2.3.1 and 2.4.1. To simulate the saturation in brightness for higher stimulation amplitudes as observed by Fernández et al. (2021, Figure 6A), this charge per second activation is passed through a sigmoidal activation curve fit on that same figure (slope$= 0.14 \cdot 10^7$, centre$=$ 2.49μC/s, see Figure 7).

(a) **(b)**

Figure 6: *Phosphenes are shown as either **(a)** a circular dot, the brightness of which is given by a Gaussian distribution or **(b)** a Gabor filtered version of the Gaussian blob, with a random orientation.*

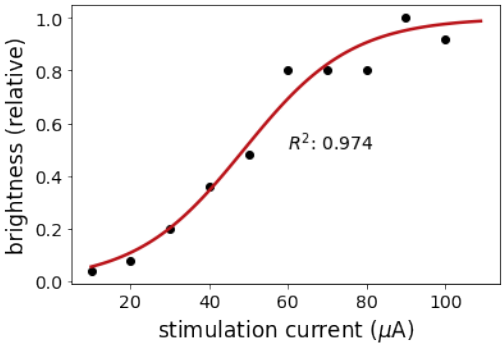


Figure 7: *Fitted sigmoidal curve describing the brightness of phosphenes for different stimulation amplitudes. Model is in red, for comparison the psychophysical data as described by Fernández et al. (2021, Figure 6a) are shown in black.*

2.7 Code and Performance

The simulator is implemented using the deep learning-library PyTorch (Paszke et al., 2019). All methods de-

scribed above are implemented as matrix operations on 3D phosphene maps (number of phosphenes x resolution width x resolution height). This allows for memory intensive but fast computations that can efficiently be executed on a GPU, making it convenient to incorporate in machine learning pipelines. Additionally, the pipeline is fully differentiable. This means that the simulator can be used as part of deep learning models (for instance in end-to-end trained image processing) or more directly for the optimisation of simulation variables, such as phosphene locations and indirectly electrode locations in the cortex.

To demonstrate the simulator, simple image-to-stimulation methods are provided. The OpenCV library (Bradski, 2000) is used to transform images to grayscale, apply gaussian smoothing and perform contour detection, either with Sobel filters or Canny edge detection. In the simulator, a sampling mask is defined for each phosphene, relating the phosphene's location in the visual field to the corresponding part of the image. The size of this sampling mask is scaled by the inverse cortical magnification factor. Each phosphene's stimulation amplitude is determined by the fraction of active pixels in the contour-detected image that overlap with the phosphene's receptive field.

To test the performance of the simulator with respect to the time it takes to generate phosphene representations, the simulator is timed while generating 100 frames of simulated phosphene vision, for different numbers of phosphenes and at varying resolutions. For this experiment, the simulator runs on a GPU (NVIDIA© Quadro RTX™ 6000) set up for parallel computing using CUDA. For each frame, a videoframe is resized to the correct resolution (256x256 pixels) and preprocessing methods as described above are applied to generate a Sobel-filtered image. The simulator's sampling mask is used to obtain stimulation amplitudes for each phosphene based on this processed image. This stimulation array is then given to the simulator, which returns a simulated phosphene image. The recorded runtimes include this entire pipeline to mirror the use case of the simulator in clinical experiments where video input needs to be transformed into phosphene vision in real-time.

2.8 Stimulation Encoding

As a showcase of a possible application of the phosphene simulator, the end-to-end framework presented by de Ruyter van Steveninck, Güçlü, et al. (2022) is adapted to use the realistic phosphene simulation. As in the original framework, the pipeline consists of an encoder, the phosphene simulator, and a decoder.

The encoder takes 256x256 images as its input, and transforms these images into stimulation patterns for N phosphenes. As in experiment 4 in de Ruyter van Steveninck, Güçlü, et al. (2022, p. 7), the encoder consists of several convolutional layers, followed by a fully connected layer which decouples the spatial information from the convolutional layer to the spatially independent phosphenes. The phosphene simulator takes these stimulation amplitudes and, using the methods explained above, generates an SPV image. Note that only the stimulation amplitudes are generated by the encoder, and the pulse width and frequency of the stimulation pulse train are assumed to be default values ($pw = 170\mu s$, $F = 200Hz$). The 256x256 SPV image is fed into the decoder, which again uses convolutional layers to form a reconstruction of the input image. The network is trained to optimise the pixel-wise mean squared error (MSE) loss between the original input image and the reconstruction. Additionally, a regularisation term is added to the cost function that pushes the encoding model toward a phosphene representation that resembles the input image.

Two experiments are conducted using our simulator. Firstly, the network is trained to reconstruct white letters against a black background. The data set and model architectures are the same as in de Ruyter van Steveninck, Güçlü, et al. (2022), and for an in-depth description we refer to that paper. The only change to the encoder was made to reflect the possibility of continuous stimulation amplitudes that our new simulator supports. To this end, the activation function of the final layer of the encoder, previously a step function to encode binary activations, is changed to a rectified linear unit (ReLU). The model is trained to optimise a combined loss function of the pixel-wise MSE between the reconstruction and the original image, and the pixel-wise MSE between the phosphene reconstruction and the original image. The simulator is initialised with 1024 phosphenes (Chen et al., 2020). To show that model performance extends to more complex images than the simple letters, the same experiment is conducted with the ADE20k data set (Zhou et al., 2019).

To demonstrate the real-life applicability of the simulator, the results of this first experiment are further extended to show the predicted outcome if stimulation values are limited to realistic ranges and safety limits. To this end, the model trained for experiment 1 is used, but during the validation stage the resulting stimulation values are capped to regions regarded safe for neurostimulation (i.e. maximum 100 μ A). To reflect that typically, neurostimulators are not capable of continuous stimulation but rather have a set of prespecified stimulation values, the stimulation predictions of the encoder are binned in 10 bins of 10 μA.

The second experiment aims to demonstrate the advantage of the temporal components of the simulator. The encoder-decoder architecture is extended to include a time dimension, allowing for the processing of video

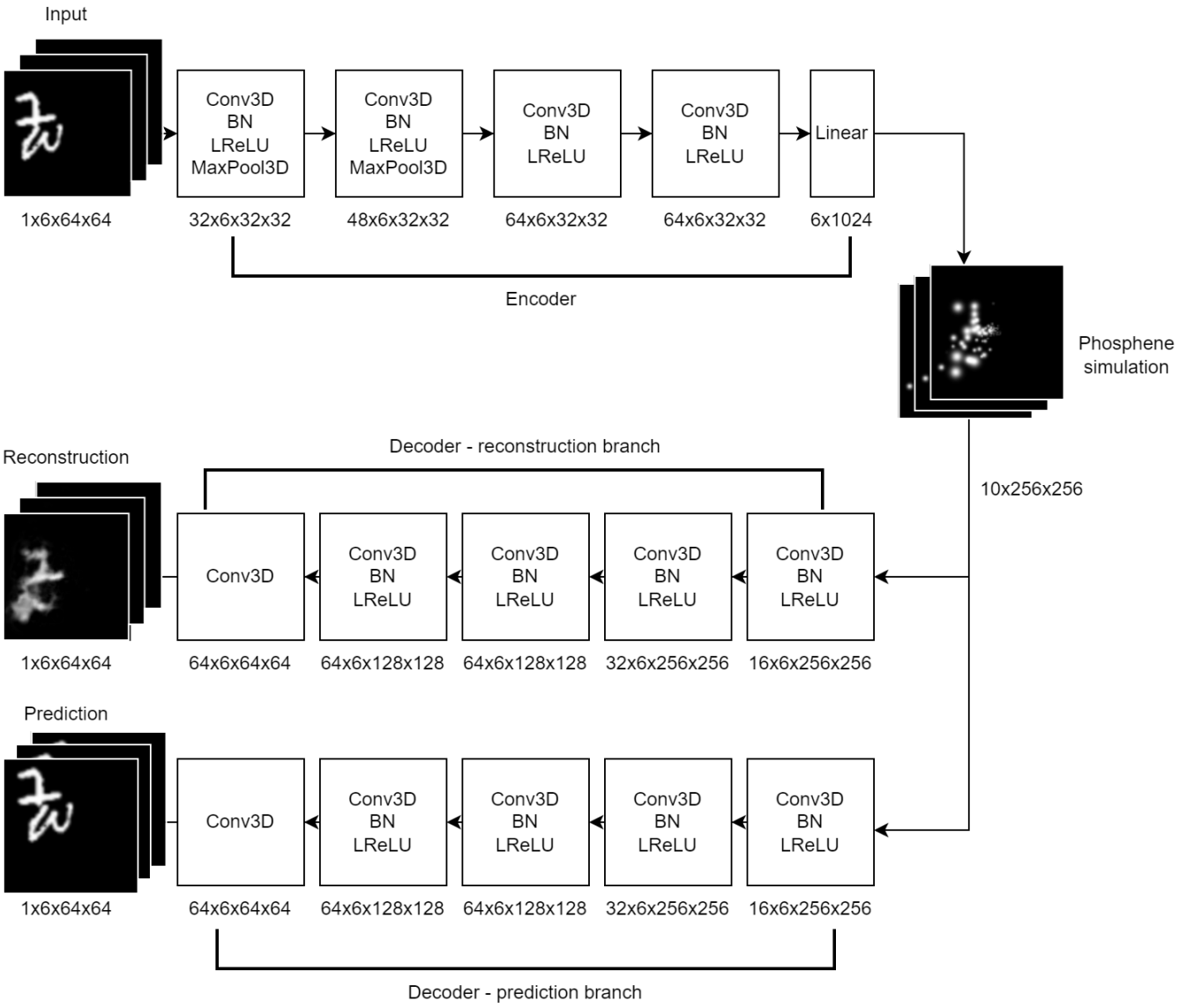


Figure 8: *Architecture of the video end-to-end pipeline. The encoder produces an n frames x m phosphenes number of stimulation amplitudes, which the phosphene simulator transforms into a phosphene representation. The two decoder branches take this phosphene representation and reconstruct the original video and predict the next n frames. n frames = 6, m phosphenes = 1024, SPV resolution = 256x256*

input. The pipeline is inspired by work by Zhao et al. (2017), where a spatio-temporal autoencoder was developed for anomaly detection in videos. To allow for feature extraction over time, the 2D-convolutions used in the previous two experiments are extended to 3D-convolutions. Additionally, the decoding part of the pipeline is duplicated, with one decoder trained to reconstruct the original video, while a second decoder is trained to predict the next few future frames. This prediction branch enables the network to learn motion features, increasing the temporal cohesion of the reconstructions and encoded stimulation patterns. The architectures of the encoder and decoder can be found in Figure 8. The performance of the model is tested on the Moving MNIST data set (N. Srivastava, Mansimov, & Salakhudinov, 2015), which contains 20-frame sequences of 64x64 images showing 2 digits moving around the frame.

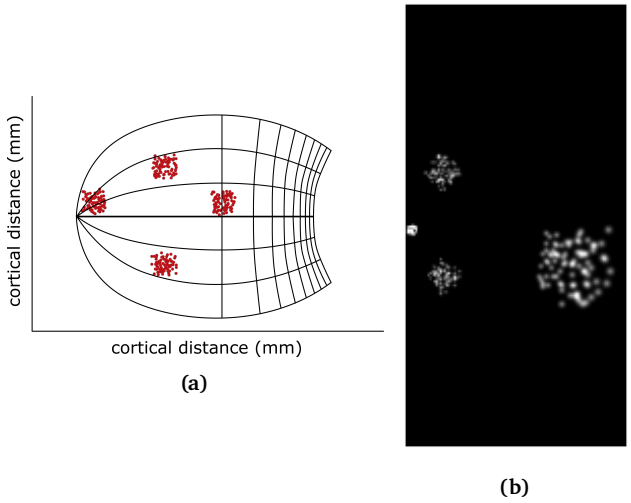


Figure 9: ***(a)*** *Electrodes placed in 4 5x5 millimetre patches on a visuotopic map (dipole model, left hemisphere).* ***(b)*** *resulting SPV image (right half visual field) for a stimulation with 80μA.*

3 Results

Figure 9 demonstrates an example result of the simulator, initialised with four 10x10 electrode patches. These electrodes' locations are specified in units of cortical distance on a dipole map of the primary visual cortex. To simulate electrode drop-out and inter-person variability, spatial noise and dropout are applied. Using the biologically grounded methods presented in the preceding sections, the resulting electrode locations are converted to phosphene locations in a map of the visual field. For a given set of stimulation parameters, for each of these phosphene the threshold, size, shape and brightness are determined, taking into account temporal effects of earlier input.

3.1 Performance of the simulator

Temporal dynamics To model the effect of stimulation over time, the simulator uses computational models of temporal dynamics. To validate the capacity of these models to accurately match psychophysical data, we fitted parameter values to behavioural data from the previously discussed literature.

The decay and increase rates for the slow trace, modelling accommodation over longer time scales, were determined by fitting the model to experimental data (Schmidt et al., 1996, Figure 5,6). Figure 10 shows the resulting effects of accommodation and recovery. This temporal model accounts for 94% of the variance present in the data (R^2=0.94).

Phosphene thresholds To validate the simulator's capacity to establish stimulation-dependent threshold values, the parameters of the thresholding model described in section 2.3.1 were fit to data presented in Fernández et al. (2021). The result of fitting Equation 9 on the 50% detection thresholds (Fernández et al., 2021, Figure 2a,b) can be found in Figure 4b, where we can see that the fitted equation accounts for 90% of the variance (R^2=0.90) for pulse widths, and 75% of the variance (R^2=0.75) for frequency.

The rheobase and chronaxie values found by fitting Equation 9 were subsequently used to generate the probability curves describing the thresholding behaviour of phosphenes. These curves can be found in Figure 11, together with a regression analysis comparing the model's predicted probabilities for different stimulation parameters to values found in psychophysical experiments. From this comparison, we can see that the model accounts for 88% of the variance (R^2=0.88).

Phosphene appearance Given the activation of a phosphene, based on the stimulation parameters, temporal effects and thresholding computations, the simulator determines the size and brightness of each of the phosphenes. Figure 7 shows the resulting relative brightness given a stimulation amplitude (pulse width and frequency are set to default values), and the data the parameters were fitted on (Fernández et al., 2021, Figure 6a). We can see that the sigmoidal curve closely matches the data (R^2=0.97). Figure 12 displays the output of the simulator for different stimulation amplitudes, illustrating the effect on size and brightness of the phosphenes.

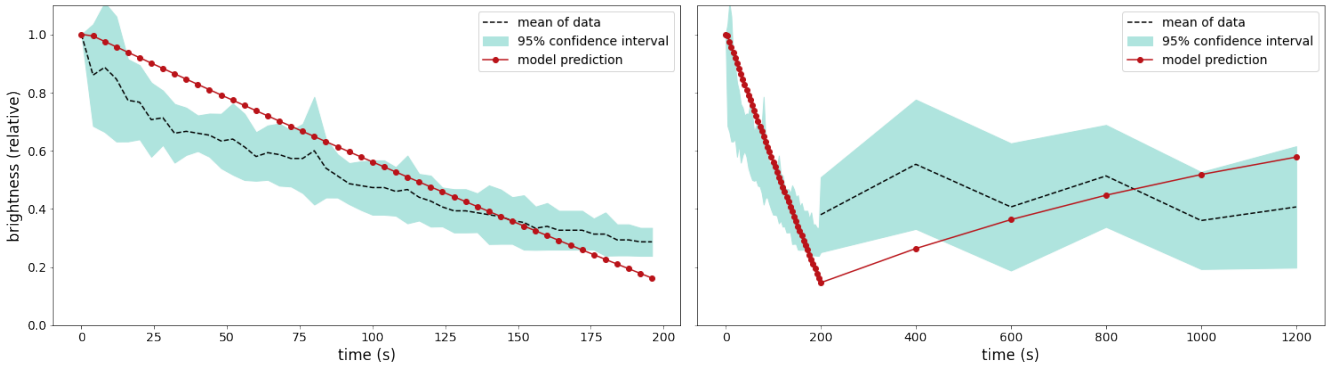


Figure 10: *Relative brightness of a phosphene in response to repeated stimulation, overlaid on experimental results by Schmidt et al. (1996). **(Left)** Relative phosphene brightness over 50 repeated stimulations at 4s intervals. **(Right)** Brightness accommodation and recovery. First 200 data points are similar to (a), the last 5 stimulations were applied every 4 minutes to test recovery.*

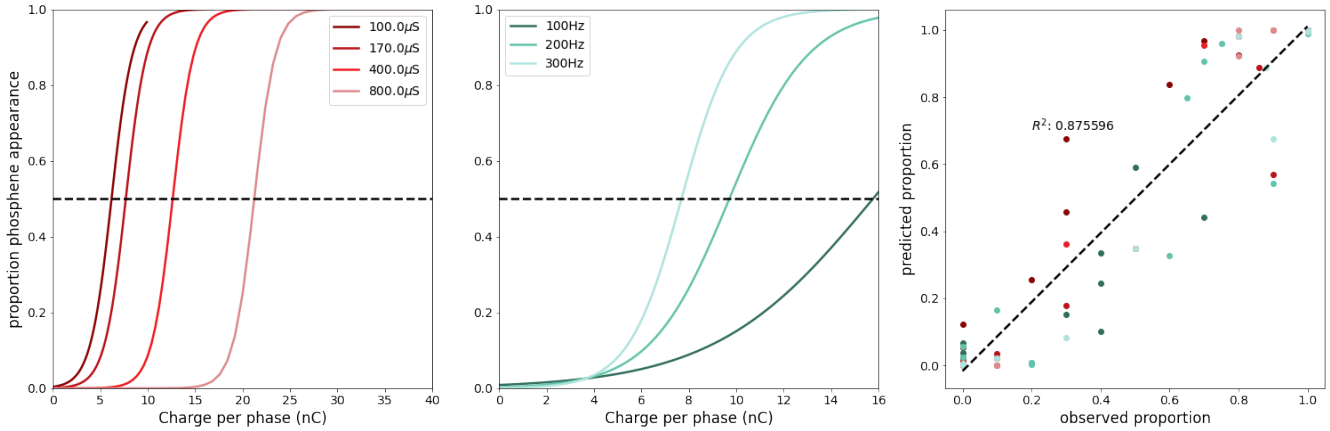
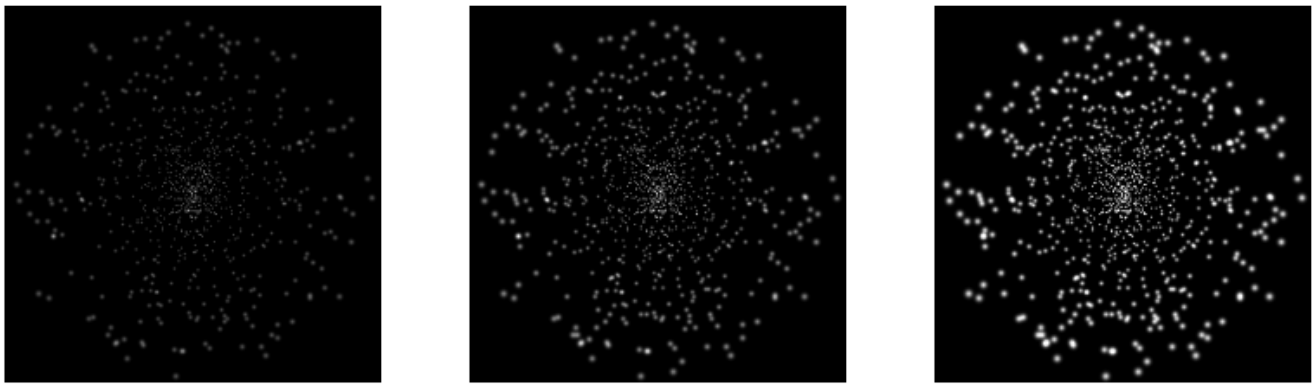


Figure 11: *The model presented in Section 2.3.1 predicts the probability of phosphenes appearing given a set of stimulation parameters. This figure shows the resulting psychometric curves for **(left)** different pulse widths, frequency is set to a default value of 300Hz, and **(middle)** different frequencies, pulse width is set to a default value of 170μs. **(Right)** Predicted probabilities compared to detection probabilities as reported in Fernández et al. (2021, Figure 2a,b). Colours conform to the conditions in the first two plots.*



(a) **(b)** **(c)**

Figure 12: *SPV-image of stimulating 952 electrodes with **(a)** 60 μA, **(B)** 80 μA, and **(c)** 100 μA stimulation current, 170μs pulse width and 300Hz frequency. 1200 electrodes were initialised on the dipole visuotopic map. After applying 20% electrode dropout, 952 phosphenes were left, distributed over a visual field of approx. 16 degrees of visual angle.*

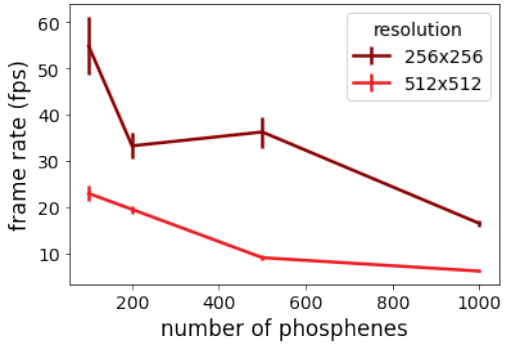


Figure 13: *Frame rate (frames per second) while processing 100 frames, for different resolutions and different numbers of phosphenes*

Run times The simulator was run on a CUDA-enabled GPU for 100 frames, for simulations with 100, 200, 500 and 1000 phosphenes, with resolutions of 256x256 and 512x512 pixels. Each setting was run five times, and the average frame rate and standard error were recorded. The resulting frame rates (in frames per second) can be seen in Figure 13. For lower phosphene numbers and a resolution of 256x256 pixels, the simulator can be run in real time, with frame rates of 35-55 fps.

3.2 Stimulation Encoding

To validate that the simulator can conveniently be incorporated in a machine learning pipeline, an end-to-end model was trained to reconstruct images of letters and naturalistic images.

For the letter-reconstruction task, the model was trained to reconstruct letters for 20 epochs. After these 20 epochs, the model showed convergence to an optimal solution (see Figure 14). The model was able to successfully reconstruct the input images, and the phosphene representations were easily interpretable as the letters in the original input. The phosphene images represented the original images to such a degree of detail that differences in font were often preserved.

Figure 14c shows the results of capping the predicted stimulation amplitudes to a maximum of 100 μA and binning the stimulation amplitudes in 10 bins of 10 μA, mirroring realistic stimulation protocols. While the reconstruction of the original image is less accurate, the letters are still recognisable in the phosphene representation.

Naturalistic images To test the model on more complex images, the same end-to-end model was trained to reconstruct naturalistic images in the ADE20K data set. The model was trained for 10 epochs, after which the loss had converged to an optimal value. Figure 15 shows

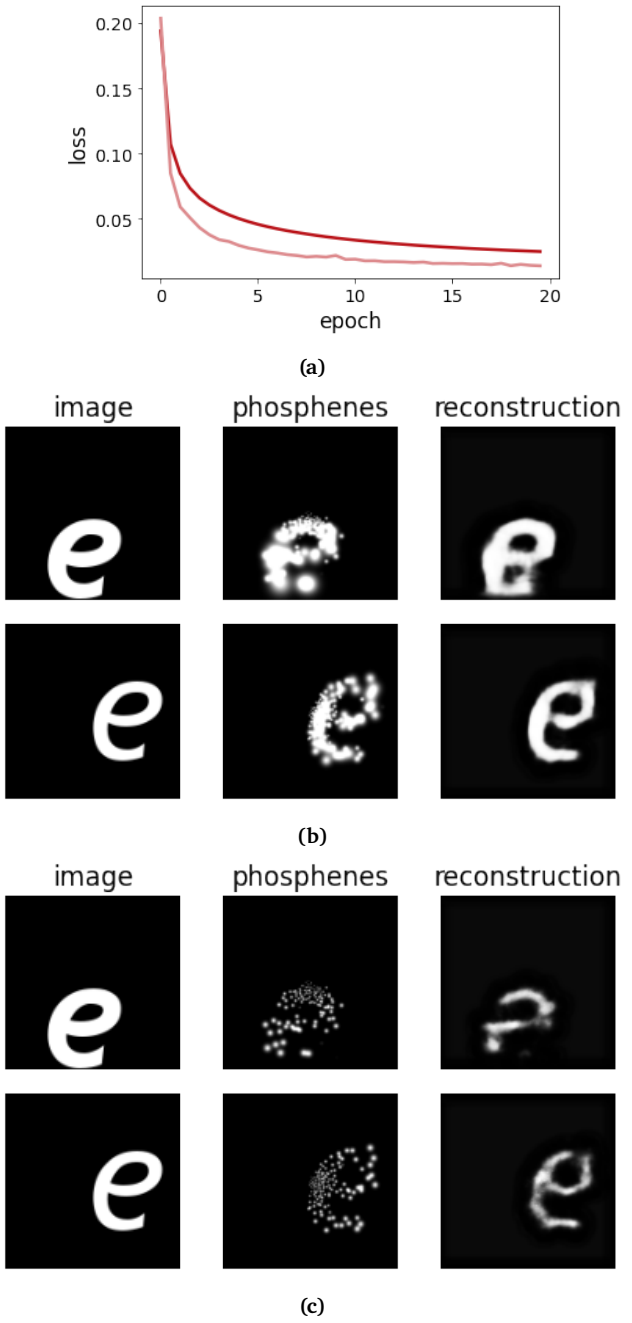


Figure 14: *Training the end-to-end model on the characters data set.* ***(a)*** *Loss curves for training and validation over 20 epochs.* ***(b)*** *Results of the trained model. Left shows the original input image, middle the phosphene representation based on the encoder's output, and right the reconstruction by the decoder.* ***(c)*** *Same trained model as (b), but stimulation values predicted by the encoder are discretised in 10μA bins and restricted to values within safety limits.*

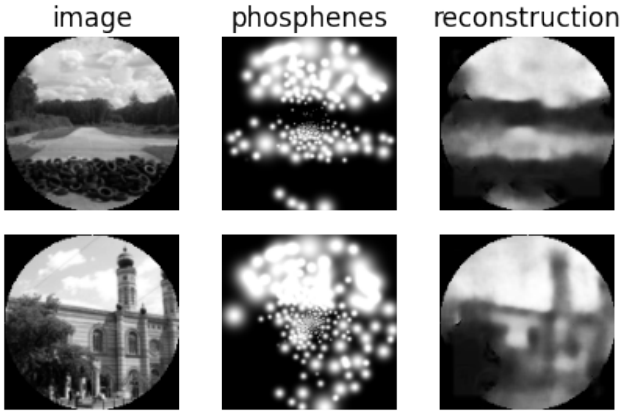


Figure 15: *Training the end-to-end model on naturalistic images in the ADE20K data set. A circular mask was applied to input images to account for the circular phosphene map. Left shows the input image, middle the phosphene representation based on the encoder's output, and right the reconstruction by the decoder.*

the resulting phosphene representations and reconstructions. The end-to-end model has learned to successfully reconstruct the original images, including grey values.

Video encoding In order to demonstrate the advantage of the temporal components of the simulator, the encoder-decoder architecture was extended to include a time dimension. This spatiotemporal model was trained for 4 epochs. Figure 16 shows several frames of the videos produced by this model. We can see that the model has learned to represent the original input frames in phosphene vision over time, and the decoder is able to reconstruct the original input approximately. For the predictions of the future frames, although the decoded images are more blurry, we can see that the model can predict the approximate direction of movement of the letters, indicating that the encoder model has learned to integrate temporal information.

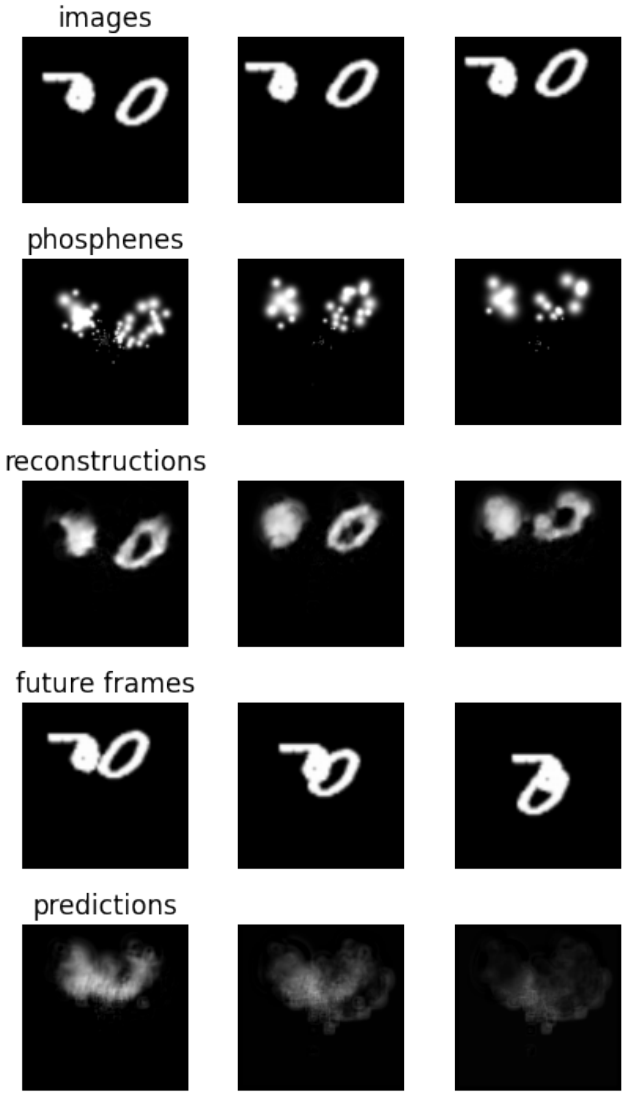


Figure 16: *Training the spatiotemporal end-to-end model on the moving MNIST data set. From left to right images in the temporal sequence are shown. The top row shows the original input frames, the second row the phosphene representation and the third row the model's reconstruction of the original input. The fourth row shows the next frames in the sequence (not provided to the encoder) and bottom row the decoder's predictions of these future frames.*

4 Discussion

This report presents a biologically plausible phosphene simulator, which takes realistic ranges of stimulation parameters, and generates a phenomenologically founded representation of phosphene vision. From the results presented above, we can see that our simulator is able to produce phosphene percepts that match the descriptions of phosphene vision that were gathered in the pioneering studies that were conducted over the past decades. Running on a GPU, the simulator is fast enough to make it a good choice for machine learning optimisation experiments. For simulations with up to 500 phosphenes,

the simulator can run in real-time, and as such could be used in clinical experiments with sighted volunteers. In combination with recent developments in Virtual Reality technology, this realistic simulator can help bridge the reality gap between simulated phosphene vision and the experience of actual prosthesis wearers.

4.1 Validity of the simulator

The results presented in the preceding section support the biological plausibility of our simulator. The mapping from electrode locations to phosphene percepts conforms to the principles of cortical magnification and the visuotopic organisation of the visual cortex, as demonstrated in Figures 9a and 12. The incorporation of the stimulation amplitude, pulse width and frequency as stimulation parameters yields effects that match descriptions in the literature: higher stimulation strengths yield larger and brighter phosphenes (e.g. Fernández et al., 2021; Schmidt et al., 1996). Higher pulse widths and frequencies yield similar effects, but these effects are adapted to mirror a loss in stimulation efficiency as found in several thresholding experiments (Niketeghad et al., 2019; Fernández et al., 2021).

Thresholding The thresholding model itself shows a good correspondence (88% of variance explained) to results reported by Fernández et al. (2021) when fit to data from that same study (see Figure 11). The model does seem to overestimate the effects of frequency, and underestimate the effects of pulse width. It should be noted that the model currently contains the most straightforward translation of the charge-duration curve to the two-dimensional pulse width and frequency domain (i.e. Equation 9). It is very possible that a function with a larger number of free parameters would be able to more accurately predict the data. However, in the light of the low amount of data available to fit the parameters on, such a simple, but relatively accurate computational model with only two free parameters can regarded to be preferable.

Temporal dynamics In our comparison with the experimental data presented in Schmidt et al. (1996), we saw that the simulator can accurately model both short and longer-term accommodation dynamics. In Figure 10 we can see that the predicted accommodation in response to repeated stimulations over the course of 200 seconds matches the observations made in accommodation experiments in that same study. However, our model predicts a moderate recovery over the next 1000 seconds, while this was not observed by Schmidt et al. Nevertheless, our predicted recovery largely stays within the 95% confidence intervals as determined by the original data. Taking into account that other studies found accommodation over different time scales and sometimes no accommodation at all (Bak et al., 1990; Bartlett et al., 1977; Dobelle & Mladejovsky, 1974; Fernández et al., 2021), we believe that the underlying model gives a plausible description of the data. Additionally, the presented model is highly flexible, and can be adapted through only five parameters to simulate a large range of possible temporal dynamics, allowing researchers to improve the simulator to better model other (future) experimental results.

Phosphene percepts The phosphenes in the simulator are shown as circular or slightly elongated white dots. Tehovnik and Slocum (2007) provide an explanation for this 'featureless' percept in terms of the spread of current as discussed in the preceding section: with the stimulation amplitudes that are used for intracortical stimulation, typically in the range of 10-100 μA, current is expected to spread to an area with a diameter of around 0.25-1 mm. Cortical columns coding for specific visual characteristics such as orientation or colour are thought to span around 0.17mm of tissue (Hubel, Wiesel, & Stryker, 1978). This means that several cortical columns are activated by each stimulating electrode, giving rise to the featureless percept. The same model predicts that lower stimulation amplitudes will only activate one cortical columns, allowing for feature-selective phosphenes. This theory seems to be supported by findings by Schmidt et al. (1996), where the participant reported seeing coloured phosphenes when an individual electrode was stimulated with $7.5\mu A$. For higher stimulation currents, that same electrode gave rise to white, yellowish or greyish percepts.

Predictability Apart from the probabilistic aspect of the thresholding model, the simulator is deterministic: phosphenes have a single, constant location, and changes in stimulation parameters will lead to predictable phosphene percepts as demonstrated in Figure 12. Clinical studies have shown that phosphene locations are stable over time (Brindley & Lewin, 1968). However, there are unpredictable and sometimes contradictory elements to phosphene vision that are not yet fully understood. Firstly, several sources note that stronger stimulation amplitudes can give rise to double phosphenes being perceived, at different locations (Brindley & Lewin, 1968; Schmidt et al., 1996; Dobelle & Mladejovsky, 1974). Secondly, simultaneous stimulation of multiple electrodes has lead to unexpected percepts. In the most recent study with intracortical electrodes implanted in a blind volunteer, researchers were able to stimulate multiple electrodes in configurations such that the participant was able to read letters. However, in some instances, the simultaneous

stimulation led to unexpected letters being perceived (Fernández et al., 2021). Lastly, while most studies report a proportional effect of stimulation strength on phosphene size, Schmidt et al. (1996) noted that for most electrodes, higher stimulation strengths led to the opposite effect of smaller phosphenes. However, this effect was not found for all electrodes.

Uncertainty in experimental data Of course, it should be taken into account that cortical neuroprostheses are still in the beginning stages of development: neurostimulation hardware and stimulation protocols are continuously being improved, and clinical trials are often limited to small numbers of blind volunteers. Thus, it is no surprise that the amount of data that is available at the present moment is limited, often open for multiple interpretations, and sometimes contains apparent contradictory information. This also manifests in the simulator presented here. The issue of contradictory data with regards to accommodation in response to cortical stimulation as described above illustrates this issue nicely.

The problem of having multiple interpretations of underlying mechanisms that explain the data is especially apparent in the thresholding mechanisms implemented in our simulator. This thresholding mechanism works on the assumption that insufficient stimulation leads to the neuronal population surrounding the electrode tip not reaching action potentials and that thus no phosphene is elicited. However, a possible different interpretation of the psychometric curves as presented in Figure 11 in this study and in Fernández et al. (2021, Figure 2) is that it is more related to the relative brightness in response to higher stimulation strengths: for insufficient stimulation, a phosphene might be elicited, but it is dim enough not to be distinguishable for the participant. Due to the subjective and behavioural nature of these clinical experiments, it can be hard to differentiate between these two possible interpretations of the data. Our model implements a version of the first explanation. It remains to be seen whether an implementation of the second interpretation would lead to other phenomenological results in practice.

In general, the points discussed above demonstrate the modular approach that was taken in the development of this simulator, and its advantage in an area of research that is experiencing such rapid developments. Most of the computational models in this simulator are parameterised by a low number of parameters that can quite easily be fit to reproduce experimental data. Both the thresholding and temporal models can be turned on or off depending on need, and different initialisation methods can be chosen to match the needs and wants of researchers. As research labs around the world are working towards a viable cortical prosthesis for the blind, it is likely that new psychophysical data will be gathered, and that computational methods can be developed to simulate these new data more accurately. Thanks to the modular approach, different parts of the simulator can be adapted separately to stay up-to-date with the state-of-the-art in cortical prosthetic developments.

4.2 Application of machine learning

The results of the computational experiments show the promise of the use of our simulator in machine learning pipelines. Using a previously proposed end-to-end architecture (de Ruyter van Steveninck, Güçlü, et al., 2022), we were able to train a neural network encoder to predict continuously-valued stimulation amplitudes. The resulting phosphene images, constructed by our simulator, were interpretable representations of the original input (Figures 14b and 15). By limiting the predicted stimulation values to safe ranges, the real-life applicability of the simulator is demonstrated. By using realistic input ranges, the simulator can be used to directly train deep learning models to predict clinically applicable stimulation protocols. The spatiotemporal end-to-end model (Figure 16) shows that these deep learning models can successfully be extended to the temporal domain.

4.3 Advantages of a realistic simulator

Using a realistic simulator of phosphene vision, as opposed to the simplified versions that are often used in SPV research, has several advantages. For instance, by incorporating visuotopic mapping in their simulations, N. R. Srivastava et al. (2009) were able to show how visual performance is impacted when the resolution of phosphene vision is higher in the central visual field than in the periphery. Ideally, using a realistic phosphene simulator closes, or at least narrows, the gap between the simulated environment and the real-life experiences of neuroprosthesis users. Results of clinical experiments with sighted volunteers, for example in the field of preprocessing camera input, are more directly applicable to the real-world situation when using a biologically plausible simulator. The fact that our simulator operates on stimulation parameters in realistic ranges enables the prototyping and evaluation of stimulation protocols by engineers. In the future, realistic simulations might aid clinicians in the development of rehabilitation and support methods for prosthesis wearers. As an example, several SPV studies already make use of simulations with multiple levels of phosphene brightness, showing the advantage of such an approach (Sanchez-Garcia et al., 2020; Vergnieux et al., 2017). However, without a realistic mapping of stimulation parameter input to

phosphene percept, it is unclear how such an encoding can be achieved practically.

Another benefit of our realistic simulator can be observed when looking at the phosphene mapping process from the reversed perspective: instead of defining electrode coordinates to determine where someone will see phosphenes, our model can also be used to determine optimal electrode locations. By defining optimal phosphene locations in the visual field (for instance to form a uniform grid-like covering of the visual field) and relating these locations back to cortical coordinates on the flattened visuotopic map, a preliminary blueprint can be made for optimal electrode locations for implantation. Due to the differentiable nature of our simulator, the optimisation of phosphene locations is a process that can be automated and implemented using state-of-the-art machine learning methods.

4.4 Future work

This biologically plausible simulator opens up a wide array of possibilities for future research. Firstly, while the present methods and parameters were developed and fit for intracortical stimulation protocols, the simulator can easily be re-calibrated to represent data from neuroprostheses that use surface electrodes.

Secondly, it would be interesting to examine whether earlier discrepancies and differences in results from SPV-studies persist when these experiments are repeated using one and the same, more realistic simulator. Such a comparative replication study would allow to determine whether the previously established lower limits for amounts of phosphenes that are needed for successful navigation or reading are still valid, or whether examined preprocessing methods still work as expected when the phosphenes are not organised in a grid-like structure. A comparative study would be able to establish an accurate baseline for performance with this realistic simulator, allowing researchers to build on existing work. Using a biologically plausible simulator would also allow for the comparison of different image preprocessing techniques such as artificial retinas, deep-learning based scene simplification or end-to-end models.

Thirdly, the use of realistically valued input stimulation values in this simulator allows for the optimisation of stimulation protocols within safety and hardware constraints. Due to limitations in the hardware and energy use of the neuroprosthetic devices, there will be constraints in the deployment of stimulation patterns. For instance, it will most likely not be possible to stimulate all electrodes in a neuroprosthetic device at the same time (Farnum & Pelled, 2020). At the same time, there are safety limits on the amount of charge and/or current that can be deposited in the brain at a given time. Using our simulator, deep-learning methods can be leveraged to produce constraint optimised stimulation paradigms.

Finally, our simulator includes a model of temporal dynamics. Future research could explore whether neural networks can be trained that leverage these dynamics to their advantage. Current preprocessing techniques often focus on frame-by-frame processing approaches. However, it as been shown that dynamic stimulation paradigms can be beneficial, both for interpretability and energy efficiency (Beauchamp et al., 2020). This realistic simulator can provide the tools to further investigate such dynamic stimulation paradigms and machine learning approaches that can learn to construct optimal stimulation protocols.

4.5 Conclusion

We have presented a biologically plausible simulator of phosphene vision. This simulator produces results that relate to findings in a wide array of experimental results from clinical studies. Its phenomenologically accurate simulations allow for the optimisation of visual cortical prosthesis in a manner that narrows the gap between simulation and reality drastically compared to previous studies of simulated phosphene vision. For low-resolution settings, can operate in real time, and is therefore a viable option for clinical experiments with sighted volunteers. Additionally, the implementation in PyTorch and its differentiable nature make it a perfect choice for machine learning approaches to phosphene vision.

References

Ackland, P., Resnikoff, S., & Bourne, R. (2017). World blindness and visual impairment: despite many successes, the problem is growing. *Community Eye Health*, *30*(100), 71–73. Retrieved from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5820628/`

Bak, M., Girvin, J. P., Hambrecht, F. T., Kufta, C. V., Loeb, G. E., & Schmidt, E. M. (1990, May). Visual sensations produced by intracortical microstimulation of the human occipital cortex. *Medical and Biological Engineering and Computing*, *28*(3), 257–259. doi: 10.1007/BF02442682

Bartlett, J. R., Robert W. Doty, S., Lee, B. B., Negrão, N., & William H. Overman, J. (1977). Deleterious Effects of Prolonged Electrical Excitation of Striate Cortex in Macaques. *Brain, Behavior and Evolution*, *14*(1-2), 46–66. doi: 10.1159/000125575

Beauchamp, M. S., Oswalt, D., Sun, P., Foster, B. L., Magnotti, J. F., Niketeghad, S., ... Yoshor, D. (2020, May). Dynamic Stimulation of Visual Cortex Pro-

duces Form Vision in Sighted and Blind Humans. *Cell*, *181*(4), 774–783.e5.

Beyeler, M., Boynton, G. M., Fine, I., & Rokem, A. (2017). pulse2percept: A Python-based simulation framework for bionic vision. *Proceedings of the 16th Python in Science Conference*, 81–88. doi: 10.25080/shinma-7f4c6e7-00c

Bollen, C. J., Güçlü, U., van Wezel, R. J. A., van Gerven, M. A., & Güçlütürk, Y. (2019, September). Simulating neuroprosthetic vision for emotion recognition. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 85–87).

Bosking, W. H., Sun, P., Ozker, M., Pei, X., Foster, B. L., Beauchamp, M. S., & Yoshor, D. (2017, July). Saturation in Phosphene Size with Increasing Current Levels Delivered to Human Visual Cortex. *Journal of Neuroscience*, *37*(30), 7188–7197. doi: 10.1523/JNEUROSCI.2896-16.2017

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Brindley, G. S., & Lewin, W. S. (1968). The sensations produced by electrical stimulation of the visual cortex. *The Journal of Physiology*, *196*(2), 479–493. doi: 10.1113/jphysiol.1968.sp008519

Bruce, I., Irlicht, L., White, M., O'Leary, S., Dynes, S., Javel, E., & Clark, G. (1999, June). A stochastic model of the electrically stimulated auditory nerve: pulse-train response. *IEEE Transactions on Biomedical Engineering*, *46*(6), 630–637.

Bruce, I., White, M., Irlicht, L., O'Leary, S., Dynes, S., Javel, E., & Clark, G. (1999, June). A stochastic model of the electrically stimulated auditory nerve: single-pulse response. *IEEE Transactions on Biomedical Engineering*, *46*(6), 617–629.

Cha, K., Horch, K. W., & Normann, R. A. (1992, July). Mobility performance with a pixelized vision system. *Vision Research*, *32*(7), 1367–1372. doi: 10.1016/0042-6989(92)90229-C

Cha, K., Horch, K. W., Normann, R. A., & Boman, D. K. (1992, May). Reading speed with a pixelized vision system. *JOSA A*, *9*(5), 673–677. doi: 10.1364/JOSAA.9.000673

Chen, X., Wang, F., Fernandez, E., & Roelfsema, P. R. (2020, December). Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. *Science*, *370*(6521), 1191–1196. doi: 10.1126/science.abd7435

Dagnelie, G. (2008). Psychophysical Evaluation for Visual Prosthesis. *Annual Review of Biomedical Engineering*, *10*(1), 339–368. doi: 10.1146/annurev.bioeng.10.061807.160529

Dagnelie, G., Keane, P., Narla, V., Yang, L., Weiland, J., & Humayun, M. (2007, February). Real and virtual mobility performance in simulated prosthetic vision. , *4*(1), S92–S101. doi: 10.1088/1741-2560/4/1/S11

Dagnelie, G., Walter, M., & Yang, L. (2006, June). Playing checkers: detection and eye–hand coordination in simulated prosthetic vision. *Journal of Modern Optics*, *53*(9), 1325–1342. doi: 10.1080/09500340600619197

de Ruyter van Steveninck, J., Güçlü, U., van Wezel, R., & van Gerven, M. (2022, February). End-to-end optimization of prosthetic vision. *Journal of Vision*, *22*(2), 20. doi: 10.1167/jov.22.2.20

de Ruyter van Steveninck, J., van Gestel, T., Koenders, P., van der Ham, G., Vereecken, F., Güçlü, U., … van Wezel, R. (2022, February). Real-world indoor mobility with simulated prosthetic vision: The benefits and feasibility of contour-based scene simplification at different phosphene resolutions. *Journal of Vision*, *22*(2), 1. doi: 10.1167/jov.22.2.1

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

Farnum, A., & Pelled, G. (2020). New Vision for Visual Prostheses. *Frontiers in Neuroscience*, *14*, 36. doi: 10.3389/fnins.2020.00036

Fehervari, T., Matsuoka, M., Okuno, H., & Yagi, T. (2010). Real-Time Simulation of Phosphene Images Evoked by Electrical Stimulation of the Visual Cortex. In K. W. Wong, B. S. U. Mendis, & A. Bouzerdoum (Eds.), *Neural Information Processing. Theory and Algorithms* (pp. 171–178). Berlin, Heidelberg: Springer.

Fernández, E., Alfaro, A., Soto-Sánchez, C., Gonzalez-Lopez, P., Lozano, A. M., Peña, S., … Normann, R. A. (2021, December). Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex. *The Journal of Clinical Investigation*, *131*(23). doi: 10.1172/JCI151331

Foroushani, A. N., Pack, C. C., & Sawan, M. (2018, February). Cortical visual prostheses: from microstimulation to functional percept. , *15*(2), 021005. doi: 10.1088/1741-2552/aaa904

Fridman, G. Y., Blair, H. T., Blaisdell, A. P., & Judy, J. W. (2010, June). Perceived intensity of somatosensory cortical electrical stimulation. *Experimental Brain Research*, *203*(3), 499–515. doi: 10.1007/s00221-010-2254-y

Geddes, L. (2004, January). Accuracy limitations of chronaxie values. *IEEE Transactions on Biomedical Engineering*, *51*(1), 176–181.

Han, N., Srivastava, S., Xu, A., Klein, D., & Beyeler, M. (2021, January). Deep Learning–Based Scene Simplification for Bionic Vision. *arXiv:2102.00297 [cs]*. Retrieved from `http://arxiv.org/abs/2102.00297`

Horton, J. C., & Hoyt, W. F. (1991, June). The Representation of the Visual Field in Human Striate Cortex: A Revision of the Classic Holmes Map. *Archives of Ophthalmology*, *109*(6), 816–824. doi: 10.1001/archopht.1991.01080060080030

Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *Journal of Comparative Neurology*, *177*(3), 361–379.

Humayun, M. S., Dorn, J. D., da Cruz, L., Dagnelie, G., Sahel, J.-A., Stanga, P. E., ... Greenberg, R. J. (2012, April). Interim Results from the International Trial of Second Sight's Visual Prosthesis. *Ophthalmology*, *119*(4), 779–788. doi: 10.1016/j.ophtha.2011.09.028

Kim, S., Callier, T., & Bensmaia, S. J. (2016, December). A computational model that predicts behavioral sensitivity to intracortical microstimulation. *Journal of Neural Engineering*, *14*(1), 016012. doi: 10.1088/1741-2552/14/1/016012

Kim, S., Callier, T., Tabot, G. A., Gaunt, R. A., Tenore, F. V., & Bensmaia, S. J. (2015, December). Behavioral assessment of sensitivity to intracortical microstimulation of primate somatosensory cortex. *Proceedings of the National Academy of Sciences*, *112*(49), 15202–15207. doi: 10.1073/pnas.1509265112

Li, W. H. (2013, December). Wearable Computer Vision Systems for a Cortical Visual Prosthesis. In *2013 IEEE International Conference on Computer Vision Workshops* (pp. 428–435).

Lozano, A., Suárez, J. S., Soto-Sánchez, C., Garrigós, J., Martínez-Alvarez, J. J., Ferrández, J. M., & Fernández, E. (2020, September). Neurolight: A Deep Learning Neural Interface for Cortical Visual Prostheses. *International Journal of Neural Systems*, *30*(09), 2050045. doi: 10.1142/S0129065720500458

Luo, Y. H.-L., & da Cruz, L. (2016, January). The Argus® II Retinal Prosthesis System. *Progress in Retinal and Eye Research*, *50*, 89–107. doi: 10.1016/j.preteyeres.2015.09.003

Niketeghad, S., Muralidharan, A., Patel, U., Dorn, J. D., Bonelli, L., Greenberg, R. J., & Pouratian, N. (2019, May). Phosphene perceptions and safety of chronic visual cortex stimulation in a blind subject. *Journal of Neurosurgery*, *132*(6), 2000–2007. doi: 10.3171/2019.3.JNS182774

Panetsos, F., Sanchez-Jimenez, A., Diaz-de Cerio, E., Diaz-Guemes, I., & Sanchez, F. (2011). Consistent Phosphenes Generated by Electrical Microstimulation of the Visual Thalamus. An Experimental Approach for Thalamic Visual Neuroprostheses. *Frontiers in Neuroscience*, *5*, 84. doi: 10.3389/fnins.2011.00084

Parikh, N., Itti, L., Humayun, M., & Weiland, J. (2013, February). Performance of visually guided tasks using simulated prosthetic vision and saliency-based cues. , *10*(2), 026017. doi: 10.1088/1741-2560/10/2/026017

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... others (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Pezaris, J. S., & Reid, R. C. (2007, May). Demonstration of artificial visual percepts generated through thalamic microstimulation. *Proceedings of the National Academy of Sciences*, *104*(18), 7670–7675. doi: 10.1073/pnas.0608563104

Polimeni, J. R., Balasubramanian, M., & Schwartz, E. L. (2006, October). Multi-area visuotopic map complexes in macaque striate and extra-striate cortex. *Vision Research*, *46*(20), 3336–3359. doi: 10.1016/j.visres.2006.03.006

Sanchez-Garcia, M., Martinez-Cantin, R., & Guerrero, J. J. (2020). Semantic and structural image segmentation for prosthetic vision. *Plos one*, *15*(1), e0227677.

Sanchez-Garcia, M., Morollon-Ruiz, R., Martinez-Cantin, R., Guerrero, J. J., & Fernandez-Jover, E. (2022). Assessing visual acuity in visual prostheses through a virtual-reality system. *arXiv preprint arXiv:2205.10395*.

Schmidt, E. M., Bak, M. J., Hambrecht, F. T., Kufta, C. V., O'Rourke, D. K., & Vallabhanath, P. (1996, April). Feasibility of a visual prosthesis for the blind based on intracortical micro stimulation of the visual cortex. *Brain*, *119*(2), 507–522. doi: 10.1093/brain/119.2.507

Schwartz, E. L. (1983). Cortical mapping and perceptual invariance: A reply to Cavanagh. *Vision Research*, *23*(8), 831–835.

Shannon, R. V. (1992). A model of safe levels for electrical stimulation. *IEEE Transactions on biomedical engineering*, *39*(4), 424–426.

Sommerhalder, J., Oueghlani, E., Bagnoud, M., Leonards, U., Safran, A. B., & Pelizzone, M. (2003, February). Simulation of artificial vision: I. Eccentric reading of isolated words, and perceptual learning. *Vision Research*, *43*(3), 269–283. doi: 10.1016/S0042-6989(02)00481-9

Sommerhalder, J., Rappaz, B., de Haller, R., Fornos, A. P.,

Safran, A. B., & Pelizzone, M. (2004, June). Simulation of artificial vision: II. Eccentric reading of full-page text and the learning of this task. *Vision Research*, *44*(14), 1693–1706. doi: 10.1016/j.visres.2004.01.017

Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). Unsupervised learning of video representations using lstms. In *International conference on machine learning* (pp. 843–852).

Srivastava, N. R., Troyk, P. R., & Dagnelie, G. (2009, May). Detection, eye–hand coordination and virtual mobility performance in simulated vision for a cortical visual prosthesis device. *Journal of Neural Engineering*, *6*(3), 035008. doi: 10.1088/1741-2560/6/3/035008

Staddon, J. E. R. (1993, March). On Rate-Sensitive Habituation. *Adaptive Behavior*, *1*(4), 421–436. doi: 10.1177/105971239300100402

Tehovnik, E. J., & Slocum, W. M. (2007, February). Phosphene induction by microstimulation of macaque V1. *Brain Research Reviews*, *53*(2), 337–343. doi: 10.1016/j.brainresrev.2006.11.001

Tehovnik, E. J., Tolias, A. S., Sultan, F., Slocum, W. M., & Logothetis, N. K. (2006, August). Direct and Indirect Activation of Cortical Neurons by Electrical Microstimulation. *Journal of Neurophysiology*, *96*(2), 512–521. doi: 10.1152/jn.00126.2006

Thompson, R. W., Jr, Barnett, G. D., Humayun, M. S., & Dagnelie, G. (2003, November). Facial Recognition Using Simulated Prosthetic Pixelized Vision. *Investigative Ophthalmology & Visual Science*, *44*(11), 5035–5042. doi: 10.1167/iovs.03-0341

Thorn, J. T., Migliorini, E., & Ghezzi, D. (2020, July). *Virtual reality simulation of epi-retinal prosthetic vision highlights the relevance of the visual angle* (Tech. Rep.). doi: 10.1101/2020.07.18.195800

Vergnieux, V., Macé, M. J.-M., & Jouffrais, C. (2014, August). Wayfinding with simulated prosthetic vision: Performance comparison with regular and structure-enhanced renderings. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2585–2588).

Vergnieux, V., Macé, M. J.-M., & Jouffrais, C. (2017). Simplification of Visual Rendering in Simulated Prosthetic Vision Facilitates Navigation. *Artificial Organs*, *41*(9), 852–861. doi: 10.1111/aor.12868

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007, October). Visual Field Maps in Human Cortex. *Neuron*, *56*(2), 366–383. doi: 10.1016/j.neuron.2007.10.012

Winawer, J., & Parvizi, J. (2016, December). Linking Electrical Stimulation of Human Primary Visual Cortex, Size of Affected Cortical Area, Neuronal Responses, and Subjective Experience. *Neuron*, *92*(6), 1213–1219. doi: 10.1016/j.neuron.2016.11.008

Zhao, Y., Deng, B., Shen, C., Liu, Y., Lu, H., & Hua, X.-S. (2017, October). Spatio-Temporal AutoEncoder for Video Anomaly Detection. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 1933–1941). Mountain View California USA: ACM. doi: 10.1145/3123266.3123451

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2019). Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, *127*(3), 302–321.