tvst

Retina

Face and Available Chair Detection and Localization With a Second-Generation (44-Channel) Suprachoroidal Retinal Prosthesis

Lisa Lombardi[1], Matthew A. Petoe[2,3], Lauren Moussallem[1], Maria Kolic[1], Elizabeth K. Baglin[1], Samuel Stefopoulos[4], Xerxes Battiwalla[4], Janine G. Walker[5], Nick Barnes[6], Carla J. Abbott[1,7,*], and Penelope J. Allen[1,7,8,*]; for the Bionics Institute and Centre for Eye Research Australia Retinal Prosthesis Consortium

[1] Centre for Eye Research Australia, Royal Victorian Eye & Ear Hospital, East Melbourne, VIC, Australia
[2] Bionics Institute, East Melbourne, VIC, Australia
[3] Medical Bionics Department, University of Melbourne, Melbourne, VIC, Australia
[4] Bionic Vision Technologies Pty Ltd, Docklands, VIC, Australia
[5] Health & Biosecurity, Commonwealth Scientific and Industrial Research Organisation, Black Mountain, ACT, Australia
[6] School of Computing, Australian National University, Canberra, ACT, Australia
[7] Ophthalmology, Department of Surgery, University of Melbourne, Melbourne, VIC, Australia
[8] Vitreoretinal Unit, Royal Victorian Eye and Ear Hospital, East Melbourne, VIC, Australia

Correspondence: Penelope J. Allen, Centre for Eye Research Australia, Level 10, 200 Victoria Parade, East Melbourne, VIC 3002, Australia.
e-mail: pjallen@melbourneretina.com.au

Received: July 25, 2024
Accepted: March 8, 2025
Published: May 8, 2025

Keywords: retinal prosthesis; bionic eye; suprachoroidal; face localization; chair localization

Citation: Lombardi L, Petoe MA, Moussallem L, Kolic M, Baglin EK, Stefopoulos S, Battiwalla X, Walker JG, Barnes N, Abbott CJ, Allen PJ. Face and available chair detection and localization with a second-generation (44-channel) suprachoroidal retinal prosthesis. Transl Vis Sci Technol. 2025;14(5):11, https://doi.org/10.1167/tvst.14.5.11

Purpose: To compare accuracy of the comprehensive vision processing (VP) algorithm (Lanczos2 [L2]) with the novel VP algorithms, face detection (FaD) and available chair detection (ChD) methods in recipients of the second-generation suprachoroidal retinal prosthesis.

Methods: Four suprachoroidal retinal prosthesis recipients (#NCT05158049) were acclimatized to new VP methods (FaD and ChD) with L2 used as control. For face localization, one or two mannequins (white/black) were forward or backward facing in three positions in a room with a white backdrop. Participants were asked to detect the number of mannequins and faces present and point to the forward-facing mannequin. For available chair localization, two mannequins (white/black) were seated in two of three chairs (white/black). Participants were asked to detect and navigate to the available chair.

Results: FaD performed significantly better than L2 for correct face detection (FaD) (81.25 $\pm$ 10.21%; L2 32.81 $\pm$ 5.98%; $P = 0.029$) and for face localization (FaD, 81.25 $\pm$ 10.21%; L2, 26.56 $\pm$ 10.67%; $P = 0.029$). The accuracy of mannequin detection was equivalent between FaD (47.22 $\pm$ 5.56%) and L2 (52.78 $\pm$ 13.98%) with one mannequin ($P = 0.457$), and with two mannequins present (FaD, 21.43 $\pm$ 18.44%; L2. 3.57 $\pm$ 7.14%; $P = 0.257$). The ChD VP method (88.89 $\pm$ 12.00% correct) performed significantly better than L2 (19.44 $\pm$ 13.22%) for localizing available chairs ($P = 0.029$).

Conclusions: FaD and ChD VP methods performed better than L2 for the purpose of localizing faces and available chairs, respectively.

Translational Relevance: New VP algorithms can improve the localization of specific object types while using the second-generation suprachoroidal retinal prosthesis.

Introduction

Retinal prostheses are designed to provide rudimentary artificial vision for people with end-stage inherited retinal disease. In recent years, gene therapy has become available for biallelic RPE65 mutation-associated Leber congenital amaurosis (voretigene neparvovec-rzyl [Luxturna]); however, this treatment is suitable for only 1 to 2% of inherited retinal diseases and is targeted toward earlier stages of disease.[1,2] In contrast, retinal prostheses are non–gene specific and can provide improved function for people with profound vision loss from several inherited retinal disease types. Several groups internationally have developed retinal prostheses in different anatomical locations including epiretinal,[3–5] subretinal,[6–9] suprachoroidal,[10–12] and transscleral[13,14] positions. Three devices (two epiretinal and one subretinal) achieved retinal regulatory approvals internationally; however, for commercial reasons they are no longer in production.[15]

The suprachoroidal approach has the advantage of a straightforward surgery, with low risks of surgical-related complications, despite being situated further from the target cells of the retina than epiretinal and subretinal devices. The Bionic Vision Technologies' second-generation retinal prosthesis has recently been implanted in four participants with end-stage retinitis pigmentosa (ClinicalTrials.gov identifier #NCT03406416) with findings of no serious adverse events and successful functional vision outcomes in laboratory and real-world environments after 2 years of follow-up.[12,16,17,18] At the end of the trial, study participants provided feedback that they would like improvements to daily tasks involving social interactions when using the device, such as the ability to detect faces and available chairs. With the current comprehensive vision processing (VP) method (Lanczos2 [L2])[19] used in the trial, although they can detect people in their environment, it remains difficult to discern if a person is facing them or not. Furthermore, it remains difficult to determine if a chair is available to sit on, particularly in settings such as a café. These specific capabilities are reported by implant recipients as being key to improving confidence and interaction in public spaces and social settings.

Novel VP (NVP) algorithms have now been developed for these specific tasks including face detection (FaD) and available chair detection (ChD). The aim of this study was to determine the efficacy of these two task-specific NVP methods compared with the comprehensive VP method, for the purposes of detecting faces and available chairs in indoor environments by recipients of a second-generation retinal prosthesis. The findings will help to determine whether the NVP methods should be incorporated into the VP software for future ongoing use by retinal prothesis users.

Methods

Participants

Four monocularly implanted suprachoroidal retinal prosthesis recipients with profound vision loss owing to retinitis pigmentosa were enrolled in a longitudinal study (ClinicalTrials.gov identifier: #NCT05158049). The study was approved by the Human Research Ethics Committee of St. Vincent's Hospital Melbourne (#220/21). Informed consent was obtained from each participant before enrolment. The study adhered to the tenets of the Declaration of Helsinki. All participants have bare light perception in both eyes and had the retinal prosthesis implanted in their worst-seeing eye.

Participant Demographics

Participant demographics are shown in Table 1. All participants completed both face and available chair localization tasks and had been using the device for 4 years in their local environments before this study.

Retinal Prosthesis and External Components

Participants were implanted previously with the Bionic Vision Technologies second-generation (44-channel) suprachoroidal retinal prosthesis in 2018 for a clinical trial of device safety and efficacy (Clinicaltrials.gov identifier #NCT03406416).[10,12,16–18,20–22] Three participants were implanted in the right eye and one implanted in the left eye, with surgical methods published elsewhere.[12,18,23] The electrode array consisted of 44 stimulating platinum electrodes (Ø = 1.0 mm) and two return platinum electrodes (Ø = 2.0 mm), on a silicone substrate (19 mm horizontally × 8 mm vertically).[10,12] Two receiver–stimulators were implanted behind the ear on the skull. These trials were conducted using an external headset (Fig. 1) worn by the participants and an external VP system run on a laboratory-based system using a laptop. This system enabled the recording of participant performance as captured by the camera and depth feed headset cameras and enabled researchers to view how the VP methods were processing the visual scene. This process also enabled VP methods to be selected as needed in preparation for each trial.

Table 1. Summary of Participant Demographics

Participant	S1	S2	S3	S4
Sex	Male	Male	Female	Male
Age at implant (years)	47	63	66	39
Age at NVP study (years)	51	67	70	43
Eye condition	RP (rod cone dystrophy)	RP (rod cone dystrophy)	RP (cone rod dystrophy)	RP (cone rod dystrophy)
Genotyping	Autosomal recessive; homozygous for deletion ADAM9	Negative for explaining phenotype	Negative for explaining phenotype; heterozygous for IFT81	Negative for explaining phenotype; heterozygous for NMNAT1; heterozygous for PEX26
Observed nystagmus	Mild	Intermittent	None	Mild
Visual acuity	Light perception both eyes	Light perception both eyes	Light perception both eyes	Light perception both eyes
ffERG stimulus light threshold (candela s/m^2)	0.1	0.1	0.001	0.001
Age when legally blind	20	34	41	13
Approximate age of useful form vision	34	43	56	19
Implanted eye	Left	Right	Right	Right

ffERG, full-field electroretinogram; RP, retinitis pigmentosa.

Figure 1. Generation 2 headset with mounted cameras for use with a left-sided 44-channel suprachoroidal retinal prosthesis (coils that connect to receiver–stimulators are not pictured). This component works in conjunction with an external processing unit or laboratory-based system.

NVP Algorithms

Trials were conducted using the NVP algorithms (FaD and ChD) compared with the comprehensive VP method (L2). The NVP used a combination of deep neural networks and analytical algorithms. The camera and depth feeds were fused to determine a depth map, which was used for the output of the algorithms. This was done using standard methods based on camera-depth calibration, which is performed as a part of the headset manufacturing. The object detection algorithms were tuned specifically for uses related to low vision. Specific phosphene mapping for these algorithms were based on different sets of phosphene activators, single phosphene (Figs. 2A and 2E), multiple phosphene (Figs. 2B and 2F), and disc activation (Figs. 2C and 2G). The single phosphene condition may be from either a single or paired electrode activation, as explained in previous publications.[17,21,24] It was necessary to use paired electrodes to produce percepts safely at locations away from the fovea because the required charge levels were higher.

NVP algorithms were implemented on a common platform for evaluating algorithms in both simulated and real-world environments. Tests with bionic eye participants were conducted using a common platform on standard computer hardware. Later, these same algorithms were ported to work on the embedded device and run at real-time or near real-time speeds.

Typical frame-rates are approximately 30 FPS, but there are occasions where some specific algorithms (namely, ChD) may run slower. In testing, the slowest observed rate was approximately 13 FPS, which was still deemed real time for participants.

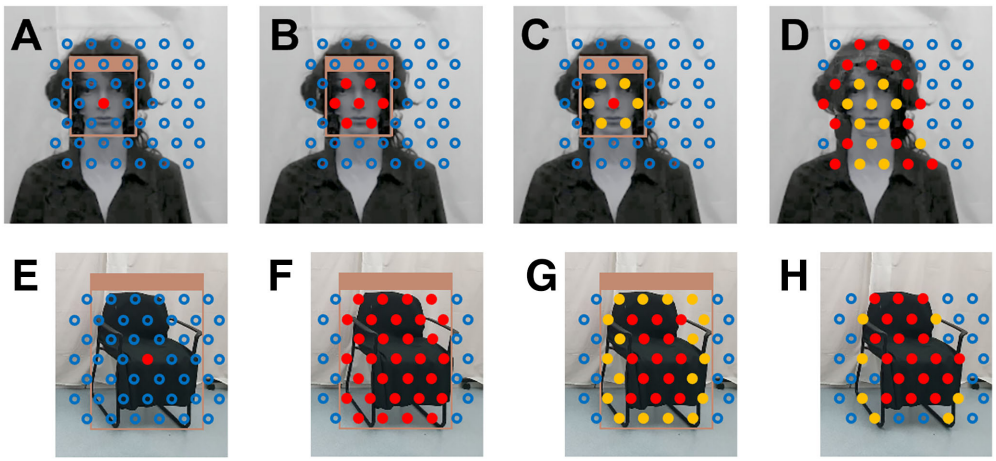

Figure 2. Indicative output for FaD algorithm: (**A**) Single phosphene. (**B**) Multiple phosphene and (**C**) Disc activation settings. ChD algorithm: (**E**) Single phosphene. (**F**) Multiple phosphene and (**G**) Disc activation settings and L2 (**D** and **H**) settings showing phosphene activation for all areas of dark contrast within the VP processing region when the camera view contains a black mannequin against a white background. Activated phosphenes are shown as *red* or *yellow circles* for all modes. *Yellow circles* indicate the weaker output for low-contrast areas with L2 or a weaker signal around the edge of the output area with FaD and ChD disc activation mode, and *red circles* indicate the stronger output. The remaining (nonactive) sampling points within the VP processing region are shown as *hollow blue circles*. Where applicable, detected faces (**A**–**C**) and available chairs (**E**–**G**) are shown within an orange boundary box. For white mannequins, L2 would be less active for low contrast areas.

Theoretical performance was evaluated against open-source validation datasets for object detectors, and internal testing using simulation. Verification testing was performed to ensure algorithms met specifications, validation testing was performed to ensure algorithm performance met expectations from simulation test results, and per-unit acceptance testing was performed for depth accuracy.

Acclimatization to Novel Algorithms

The two novel algorithms (FaD and ChD) were incorporated into the laboratory-based processing system for use. The participants were acclimatized to the NVP methods during laboratory-based sessions separate to the data collection sessions. The acclimatization process introduced the NVP methods and facilitated understanding of how they worked compared with the comprehensive, intensity-based VP method (L2). L2 is a bandlimited downsampling filter that results in a gradual fade-in or fade-out when scanning the boundaries of high-contrast objects.[19] This was configured to inverted dark mode throughout the experiment, providing phosphene output in the presence of high contrast (black objects against a white background) and little or no phosphene output for low contrast (white objects against a white background) (Figs. 2D and 2H). FaD and ChD are implementations of computer-based image recognition and were configured only to provide phosphene feedback in the presence of a face (forward facing mannequin) or an available chair, respectively.

FaD and ChD NVP algorithms are available with different settings, which produce a different user experience according to the intersection of the camera sampling points and the target object. Moreover, FaD and ChD used a depth-sensing feature of the stereo camera headset (Fig. 2) and, in contrast with L2, were calibrated to produce maximal output at close proximity to the target and to attenuate output with increasing distance from the target up to a maximum of 4 m (zero output). The three settings were developed in a codesign approach with input obtained from our participants.

During acclimatization, participants had the opportunity to evaluate the three different settings, verbally describe their experience, and then select their preferred setting for use, based on their visual experience and comfort during the trials. The three settings available for both FaD and available ChD were single phosphene, multiple phosphenes, and disc activation.

1. Single phosphene: Participants visually experience a single phosphene (which may be a result of a single or paired electrode activation)[17,21,24] closest to the center of the face (*red circle*, Fig. 2A) or available chair (Fig. 2E) when the target is within the camera view. In this mode, the phosphenes may appear brighter when a face or chair is closer in proximity, but it does not become larger.

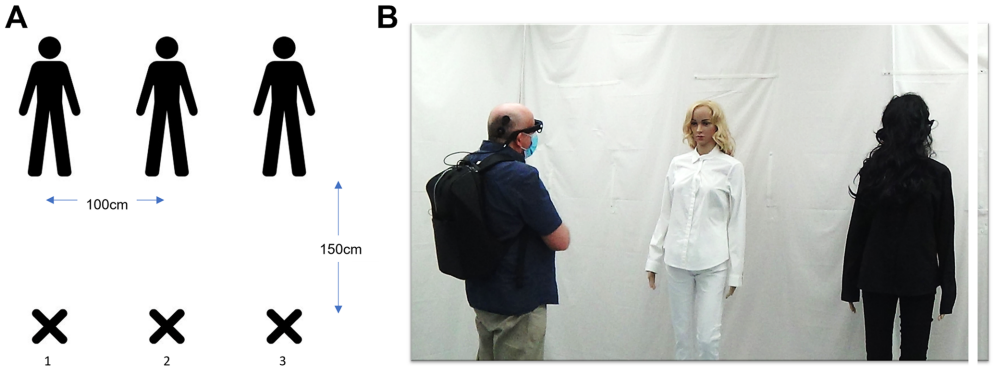

Figure 3. Schematic for FaD task set up. (**A**) The randomized stationary starting position (1 = left, 2 = middle, 3 = right) for participants denoted by the black X. The illustrated figure represents the possible mannequin positions, which will be either forward or backward facing. The distance between the possible mannequin positions (*left*, *middle*, *right*) was 100 cm and the fixed distance between the row of mannequins and the row of starting positions was 150 cm. The number of mannequins present was always 1 or 2, whereas the number of faces ranged between 0 and 2. (**B**) A participant completing a face localization trial in the laboratory.

2. Multiple phosphenes: Participants visually experience one or more phosphenes (which may be a result of one or more electrode activations). All electrodes corresponding to the intersection of the camera sampling points and the face (*red circles*, Fig. 2B) or available chair (Fig. 2F) are activated. If the object covers a region of the participant's field of view corresponding with more than one electrode, all electrodes with a center in the region will be activated, which results in multiple phosphene experiences. In this mode, the phosphenes may appear brighter and more of them may be activated when the target object is closer in proximity.
3. Disc activation: Participants visually experience one or more phosphenes (which may be activated by one or more electrodes). The intensity of phosphene experienced (and the current through the electrodes) diminishes the further away from the center of the object. All electrodes corresponding with the disc around the center of the face (*red circle*, Fig. 2C) or chair (Fig. 2G) are activated. The further the phosphene is from the center of the disc, the lower the intensity (*yellow circles*), and vice versa. In this mode, the phosphenes may appear brighter and more of them may be activated when the target object is closer in proximity.

Acclimatization session times ranged from 23 to 74 minutes (for the FaD VP method) and 30 to 48 minutes (for the ChD VP method). The session times were not restricted, instead allowing participants time to explore the NVP, become familiar with their use, and select their preferred setting for each VP method.

At the next participant sessions, before the commencement of the face and available chair localization trials, participants had training (≤60 minutes) on the tasks to be conducted for the study. Training included familiarization and orientation of the test space and a practice run of the assessment using the task-related objects (mannequins and chairs).

FaD and Localization Trials

One or two mannequins dressed in white clothes with blonde hair or black clothes with black hair were positioned either forward facing (face visible) or backward facing (face not visible) in the left, middle, and/or right positions in a square 4 × 4 m room with a white curtain backdrop (Figs. 3A, 3B). The room luminance ranged between 163–190 lux. There were three possible stationary starting positions denoted by the black X in Figure 3A where 1 = left, 2 = middle and 3 = right. The row of starting positions was 150 cm away from the row of mannequins. Participants were guided to each starting position, but were not verbally advised of the position they were in. The three possible positions for the mannequins were 100 cm apart. Selection of object and location from the pool of 108 potential combinations was randomized by computer (www.random.org), balanced for the VP method (L2 or FaD), the number of mannequins present, and the choice of black or white mannequin. There were no time constraints imposed on participants to complete each trial.

Forty randomized and timed trials were conducted, each involving FaD or L2. Some trials were discarded owing to procedural error (e.g., incorrect device mode selected), leaving 32 trials per participant with identical counterparts across participants and device modes.

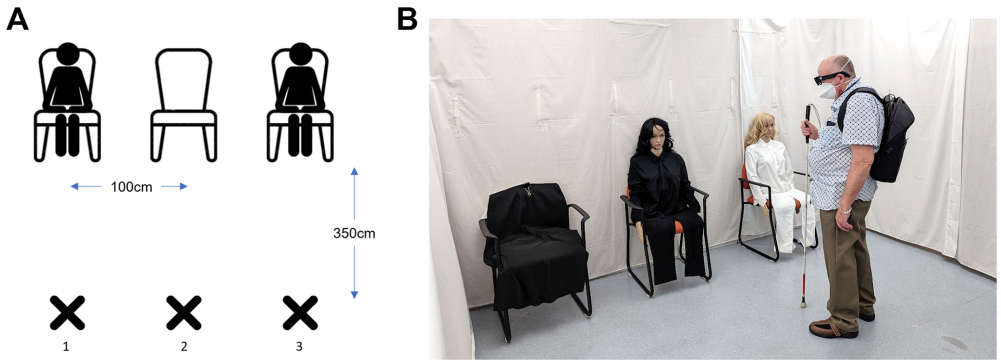

Figure 4. (**A**) Schematic for available ChD task set up. The randomized stationary starting positions (1 = left, 2 = middle, 3 = right) for participants denoted by the black X. The mannequins occupied two of three available chairs, leaving one available black or white chair. The distance between the chairs, from a centered marker was 100 cm and the fixed distance between the row of chairs and the row of starting positions was 350 cm. (**B**) A participant completing an available chair localization trial in the laboratory with mannequins seated in position.

A standardized script was delivered to each participant, which did not disclose the potential number of mannequins or faces present (because there was only ever a maximum of two mannequins or two faces present). Each trial was commenced with a verbal prompt and participants were free to move if they chose to. Participants were blinded to the VP method in use. Participants were asked to scan the scene and call out "stop" once they had decided how many mannequins were present. Early guesses were not permitted before the participant announcing "stop." They were then asked to state the number of mannequins present, state the number of faces detected, and then point to localize which of the present mannequins were forward facing. The data collected included the number of mannequins present, the number of faces reported, the pointing position (left, middle, or right) and the time taken to complete each trial.

Success at detecting the number of faces and mannequins per trial was counted when the stated number of faces and mannequins present in each trial matched the actual number of faces and mannequins present. Success at pointing to the location of a face was considered when the participant accurately pointed to the location of a forward-facing mannequin (left, middle, or right) and was considered inaccurate if they pointed to an empty space or backward facing mannequin.

ChD and Localization Trials

Two mannequins dressed in white clothes with blonde hair or black clothes with black hair were seated facing forward in two of three chairs (left, middle, and/or right) in a square 4 × 4 m room with a white curtain backdrop (Figs. 4A, 4B). The room luminance ranged between 176 and 184 lux. The unoccupied (available) chairs were either white or black. Participants started in one of three possible stationary starting positions, denoted by the black X in Figure 4A, where 1 = left, 2 = middle and 3 = right, at a distance of 350 cm away from the row of chairs. The chairs were each 100 cm apart. Participants were guided to each starting position, but where not verbally advised of the position they were in. Thirty-six trial sequences were randomized from 72 possible combinations (random.org), balanced for VP method (L2 or ChD), available chair color, and the angle between participant starting position and target location. The 36 omitted trials were duplications of chair positions that were occupied alternately by 1 white and 1 black mannequin vs. 1 black and 1 white. There were no time constraints imposed on participants to complete each trial.

Each trial was timed and commenced with a verbal prompt, after which participants were free to move if they chose to. Participants were blinded to the VP method in use. Participants were asked to scan the scene, detect the available chair, navigate with their cane toward it, and stand in front of it without making contact. Once the participant had determined their answer, they were asked to say "finish" and provide their response. Early guesses were not permitted before the participant announcing "finish."

Success at detecting an available chair occurred when the participant was able to successfully point to the location of an unoccupied chair and not point to any occupied chairs or in any other incorrect direction.

Statistical Analysis

For both the face and chair localization trials, data (responses to each trial including mannequin detection, FaD, pointing accuracy, and ChD) for each

participant were averaged. Participant averages were then compared across the two VP methods (FaD and L2 for face trials and ChD and L2 for available chair trials) using a median difference permutation test (function 'percentileTest') in the R programming language.[25,26] Each of 1 million permutations computed the difference in medians between a random reassignment of data (to each VP method) vs. the difference in medians in the original data. This approach makes minimal assumptions about the underlying data distribution and is robust for the small number of participants in this dataset. P values were multiplied by the number of groupwise comparisons to maintain a Bonferroni-corrected alpha value of $\alpha = 0.05$.

Results

Acclimatization Setting Selection for NVP Algorithms

The NVP algorithms that were selected by each participant after acclimatization are outlined in Table 2. All four participants preferred the multiple phosphenes setting when using FaD; however, preferences ranged across all options for ChD.

After acclimatization, participants reported that they felt it was more useful than their existing VP for face and available chair detection. They all felt it would be useful for social gatherings, at the dinner table, and in meetings to improve social interactions. Comments were collected regarding how these algorithms would help them with locating people more precisely and engaging with them, increase independence, enable more control over interactions, and be able to look at someone in the face rather than elsewhere.

Table 2. Preferred NVP Algorithm Settings as Selected After Acclimatization

	NVP Algorithm Setting	
Participant	FaD	ChD
S1	Multiple	Single
S2	Multiple	Disc activation
S3	Multiple	Multiple
S4	Multiple	Multiple

FaD and Localization Task

For the task of verbalizing the number of mannequins detected per trial, there was no statistical difference between the performance of FaD (35.94 ± 10.67%) and L2 (31.25 ± 8.84%; $P = 0.686$) (Fig. 5A). Upon separating the analysis by the number of mannequins presented, there was still no significant difference between VP methods for trials involving one mannequin (FaD, 47.22 ± 5.56%; L2, 52.78 ± 13.98%; $P = 0.457$) or two mannequins (FaD, 21.43 ± 18.44%; L2, 3.57 ± 7.14%; $P = 0.257$) (Fig. 5B). Further to this, detection scores were analyzed in isolation for single mannequin trials, to determine any source of response bias. When a single black mannequin (high contrast) was presented, mannequin detection scores were better (although not significant) when using L2 (95.00 ± 10.00%) vs. FaD (40.00 ± 28.28%; $P = 0.057$). In contrast, when a single white mannequin

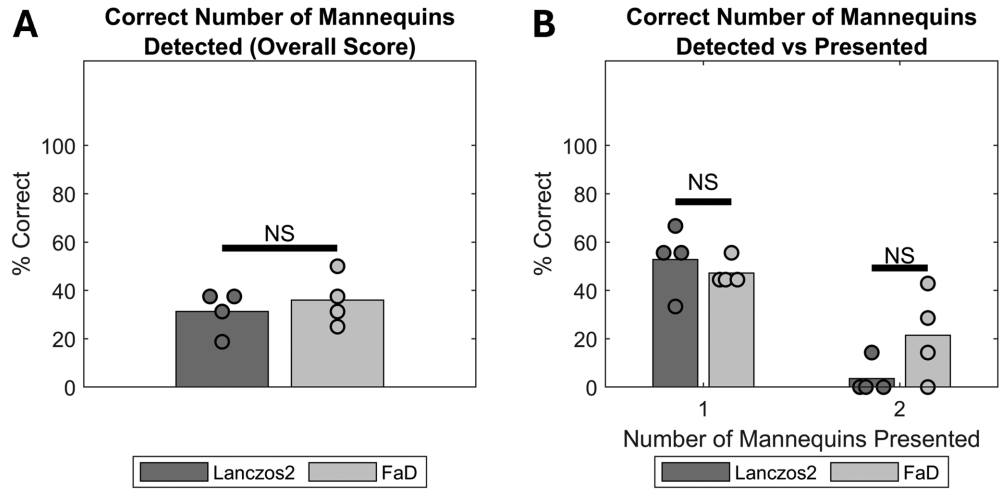

Figure 5. Correct number of mannequins identified per trial. (**A**) Overall score (%). (**B**) Score (%) separated into trials with one or two mannequins presented. Overall, there is no difference between methods for detecting the number of mannequins presented. NS, not significant.

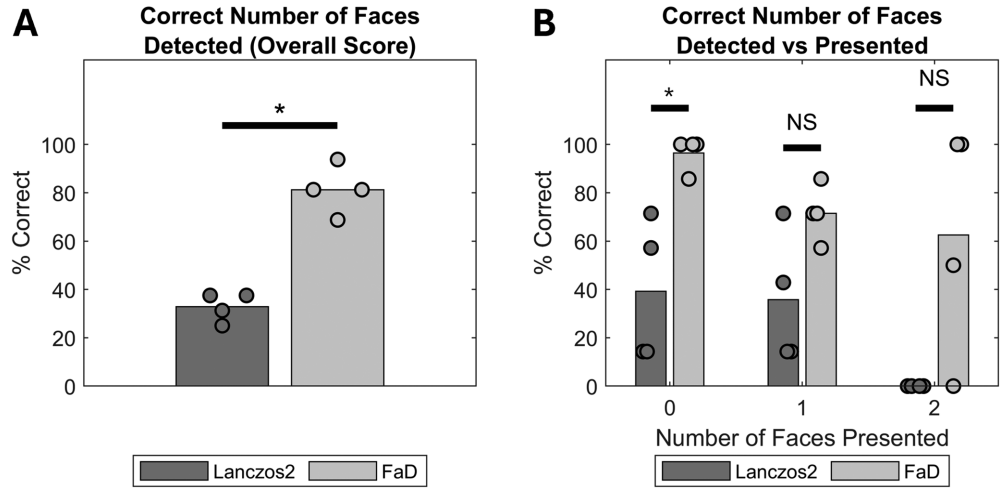

Figure 6. FaD trials overall results. (**A**) Correct number of faces detected (overall score, %) using L2 VP and Faces NVP. (**B**) Correct number of faces detected vs. the number presented (%). Faces NVP performed significantly better than L2 VP overall. *$P < 0.05$. NS, not significant.

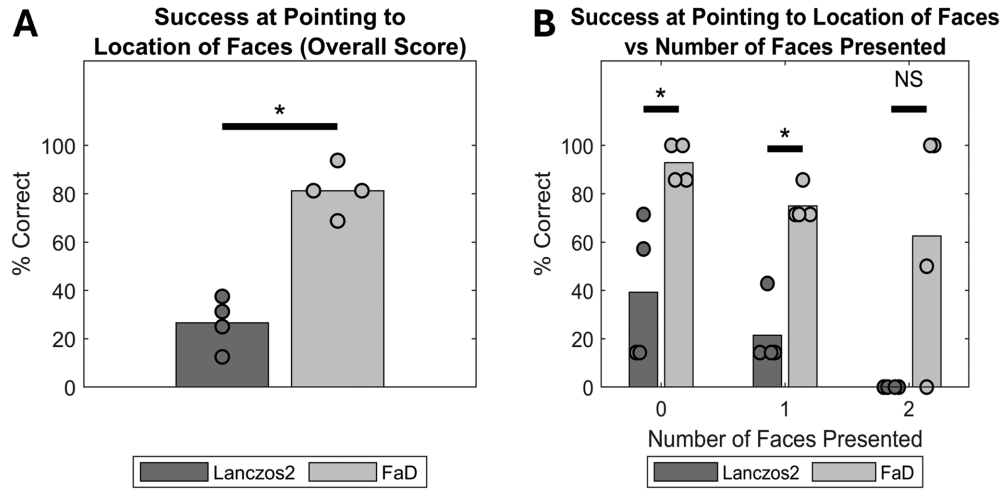

Figure 7. Success at accurately pointing to the location of a face. (**A**) Overall score (%) using L2 VP and FaD NVP. (**B**) Score (%) by the number of faces presented wherein a correct response for zero faces required that the participant suppress any response. FaD NVP performed significantly better than L2 VP overall and for when zero or one faces were presented. *$P < 0.05$. NS, not significant.

(low contrast) was presented, mannequin detection scores were significantly better when using FaD (68.75 ± 12.50%) vs. L2 (6.25 ± 12.50%; $P = 0.028$).

For the task of verbalizing the number of faces detected per trial, a mean of 81.25 ± 10.21% of presented faces were detected correctly with FaD, which was significantly better than 32.81 ± 5.98% of faces with L2 ($P = 0.029$) (Fig. 6A). When the analysis was separated by the number of faces presented, participants were significantly better with FaD at detecting when zero faces were present (FaD, 96.43 ± 7.14%; L2 39.29 ± 29.45%; $P = 0.029$). FaD also seemed to perform better than L2 for the other scenarios, but did not reach significance for the subset of one face (FaD, 71.43 ± 11.66%; L2, 35.71 ± 27.36%; $P = 0.114$) or two face (FaD, 62.50 ± 47.87%; L2, 0.00 ± 0.00%; $P = 0.143$) presentations (Fig. 6B).

For the task of localizing by pointing to each of the faces present, the FaD NVP was significantly more accurate (correct 81.25 ± 10.21% of the time) than the L2 VP (correct 26.56 ± 10.67% of the time; $P = 0.029$) (Fig. 7A). FaD also performed significantly better when assessing pointing responses in the case of trials involving zero faces (FaD, 92.86 ± 8.25%; L2, 39.29 ± 29.45%; $P = 0.029$) and one face (FaD, 75.00 ± 7.14%; L2, 21.43 ± 14.29%; $P = 0.029$), but did not reach significance for two faces (FaD, 62.50 ± 47.87%; L2, 0 ± 0%; $P = 0.143$) (Fig. 7B).

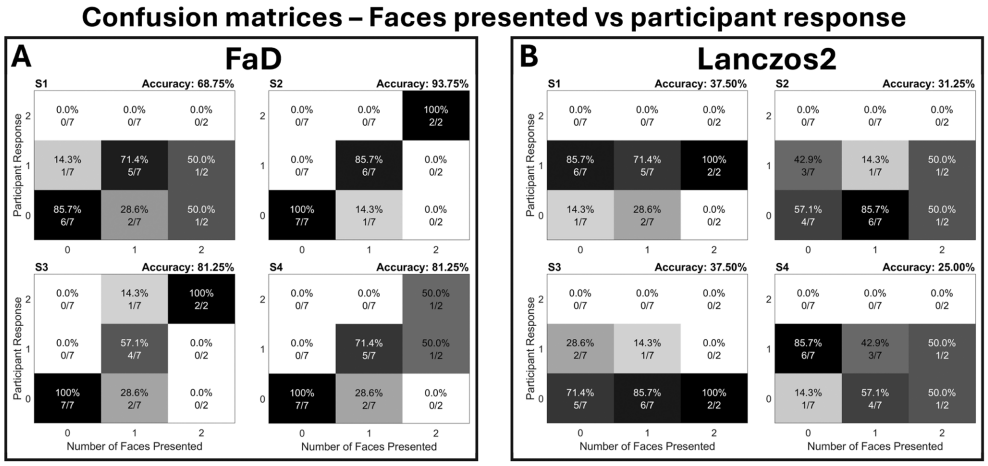

Figure 8. Confusion matrices illustrating accuracy of FaD for S1, S2, S3, and S4. (**A**) FaD. (**B**) L2 VP trials. The number of faces present (x-axis) compared with the participant response (*y* axis) are plotted. Each cell contains a percentage and ratio indicating the detection accuracy for the corresponding scenario. The numerators indicate the number of times that participant response was given and the denominator is the number of times that number of faces was presented.

The rate of false positives that occurred for the detection of faces was also analyzed by participant across all trials involving zero and one faces. The trials involving two faces cannot be used to analyze false-positive rate of detection because participants never responded that there were more than two faces present. Fewer false positives (where the participant response was higher than the number of faces presented) occurred with FaD (Fig. 8A) in comparison with L2 (Fig. 8B). The rate of false positives using FaD ranged from 0% to 7.14% (overall average across all participants, 3.57%), with S2 and S4 responding with no false positives across all FaD trials. In contrast, the false-positive rate for L2 ranged from 14.29% to 42.86% (overall average across all participants, 30.36%), with all four participants never responding that there were two faces present for the L2 trials.

Where there were no faces present (0 face trials), FaD demonstrated a very high rate of accuracy (85.7%–100%; overall, 96.43%), compared with L2 (14.3%–71.4%; overall, 39.29%). Where the participant response exactly matched the number of faces present, FaD again enabled a more accurate result ranging from 68.75% to 93.75% (overall average, 81.25%) compared with L2 (ranging from 25.00% to 37.50%; overall average, 32.81%), which is consistent with Figure 6.

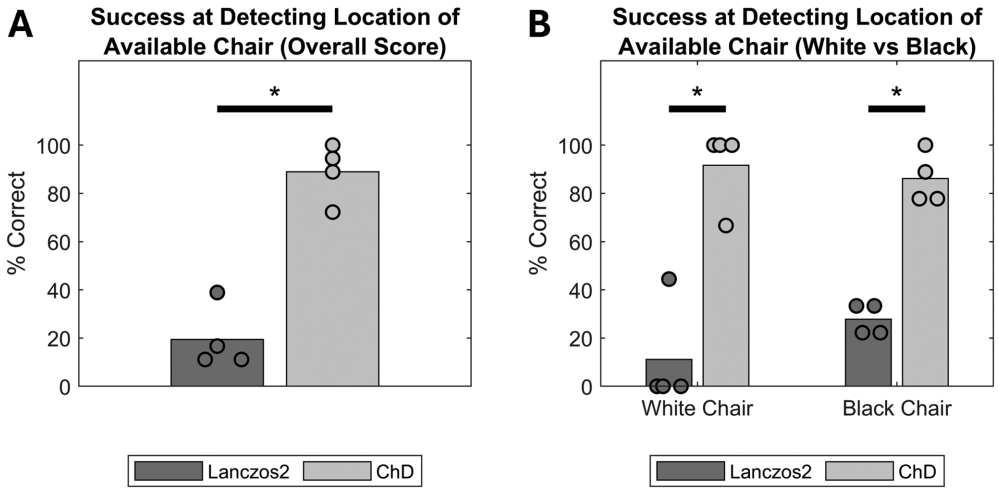

Figure 9. Correct localization of available chair using L2 VP and ChD NVP. (**A**) Overall score (%). (**B**) Separated into white and black chairs (%). ChD scored better than L2 VP in each scenario, $^*P < 0.05$.

False negatives (where the participant response was lower than the number of faces presented) can also be ascertained by reviewing the confusion matrices in Figure 8. FaD continues to show a higher rate of accuracy, via lower false-negative values, which ranged from 11.11% to 44.44% (overall average, 27.78%), compared with L2, which ranged from 44.44% to 88.89% (overall average, 72.22%). Overall, the time taken to complete each trial was not significantly different between the two VP methods (FaD, 38.77 ± 14.29 seconds; L2, 41.61 ± 16.14 seconds; $P = 0.800$).

Available Chair Localization Task

The ChD NVP method performed better than the L2 VP method for detecting and localizing available chairs. Overall, participants correctly identified 88.89 ± 12.00% of available chair presentations with ChD, compared with 19.44 ± 13.22% when using L2 ($P = 0.029$) (Fig. 9A). The ChD method performed significantly better than L2, irrespective of whether a white chair (ChD, 91.67 ± 16.67%; L2, 11.11 ± 22.22%; $P = 0.029$) or black chair (ChD, 86.11 ± 10.64%; L2, 27.78 ± 6.42%; $P = 0.029$) was present (Fig. 9B). There was no significant difference in the time taken to detect, navigate, and localize the available chair between the two VP methods; L2 took an average of 51.69 ± 25.11 seconds, whereas ChD took 48.46 ± 28.71 seconds ($P = 0.829$).

Discussion

This study shows successful use of NVP methods by participants with a suprachoroidal retinal prosthesis to perform specific tasks with the aim to improve social interactions. In comparison to a comprehensive, intensity-based VP method (L2), the results with NVP show improved accuracy of both face and available ChD and localization in an indoor environment, with both low-contrast and high-contrast objects.

The NVP methods in the present study were developed on the basis of feedback from research participants with lived experience using a suprachoroidal retinal prosthesis in their local environments. Although object detection using the original VP method (L2) was demonstrated in our earlier clinical trial,[12,24] participants noted their ability to identify objects was challenging without additional sensory input. This factor adversely affected real-world tasks involving social interactions, such as face and available ChD. The next generation of NVP methods were designed to provide users with an improved confidence in face and available chair detection localization in a way that could ultimately be used as adjunct tools to the comprehensive VP method (L2).

To ensure reliability under different scenarios of use, both algorithms were validated against open-source validation sets and performance was confirmed against precision, recall, and intersection over union metrics. Furthermore, simulation testing of the new VP methods was performed to ensure accuracy within the range tested before use with the participants. These approaches have ensured high algorithm reliability. Importantly, FaD and ChD processing methods only provide phosphene stimulus to the participant in the presence of the specific objects they are programmed to detect. This strategy enables increased certainty around identification of the object detected. For example, in FaD mode, the participant can scan the environment until a face is located. Similar to previous pilot studies with the Argus II retinal prosthesis (Stanga et al., IOVS, 2013, 54, ARVO E-Abstract, #1766), once the software detects the presence of a face in the camera view of the headset worn by the participant, the user will experience phosphene vision indicating that their head and camera angles are in the direction of a face. The same occurs with ChD mode, which will only provide a phosphene stimulus when the headset camera has been aligned with an available chair. If the chair is occupied, no stimulus is provided and the user must continue to scan the environment looking for an available chair.

This simplification of the visual input provided to participants, made possible by the NVP algorithms, enables a high level of confidence that a face or available chair has been identified. Similarly, others have attempted to simplify the information provided to Argus II users by incorporating a heat-sensitive camera to assist with tasks that are temperature sensitive.[27] Although these options provide increased certainty around object detection, they are best used in conjunction with other algorithms to ensure visual inputs are not missed for objects these modes are not attuned to detect.

The use of computer vision algorithms to improve VP for prosthetic devices may also enable improved understanding of surrounds, but at this stage has only been shown to improve this for sighted participants viewing phosphene simulations.[28,29] White and black objects were selected for FaD and ChD trials (mannequins dressed in white clothes with blonde hair or black clothes with black hair and white and black chairs, respectively) against a white background to represent high- and low-contrast scenes and reflect a real-world indoor environment. L2 (dark mode) is an intensity-based VP that relies on the presence of a darker object against a lighter background to

achieve detection of an object.[19] This means that L2 works well in high-contrast settings and not as well in low-contrast settings. Furthermore, the L2 method can confuse shadows for objects or not work well in low light levels or when viewing the same scene from a different angle. Hence, as expected, participants found it difficult to detect low-contrast white chairs against a white background with L2. However, they could detect both black and white available chairs (high- and low-contrast objects) with ChD. Furthermore, even in the high-contrast setting (black chairs against white background), ChD was much better at correctly detecting available black chairs than L2 (Fig. 9B). This is owing to the specific design of the ChD NVP method for the purpose of available chair detection only, whereas the L2 VP method will provide phosphene stimulus for any dark object, irrespective of whether the object is an available or unavailable chair.

The use of FaD and ChD NVP methods by retinal prosthesis users has demonstrated improved task-specific object detection, identification, and localization for faces and available chairs, irrespective of the object contrast with the surrounding environment. Analysis of the time taken to complete trials further consolidates this improved functionality; no additional time was required to complete trials using FaD NVP, further validating its outperformance of L2 for these purposes. However, neither L2 VP nor FaD NVP were effective when detecting the number of mannequins present (which included mannequins facing away), with less than 40% of mannequins detected by each method (Fig. 5). An improved general purpose mode, which is less dependent on contrast or the presence of a face in the scene, is required to improve the accurate detection of people, irrespective of the direction they are facing and whether they are in high or low contrast environments. This is a potential area of future research.

The accuracy with which FaD NVP functions was demonstrated by reviewing the rate of false-positive identification of faces using this method (Fig. 8A) compared with L2 (Fig. 8B). The optimal scenario is achieved when the plot illustrates a diagonal result (i.e., cells darkest from bottom left to top right), indicating good correlation between the participant's response and the number of faces present per trial.

Each of the four participant confusion matrices for FaD depict a diagonal trend, indicating that participant responses tended to reflect the true number of faces present in each trial. For FaD, both the rate of false positives and the rate of false negatives were lower compared with L2. This result suggests a higher rate of accurate face detection when using FaD NVP and confirms the benefit of this method compared with L2. However, it should be noted that there were a low number of trials involving two faces (two each for FaD and L2), which form part of the analysis of false-negative responses. Scenarios involving two faces and responses of zero or one faces and trials involving one face with a response of zero faces were included when considering the false negative results. If anything, this assists the L2 overall detection and localization result, because this VP does not detect low-contrast objects well. Despite this, FaD has still outperformed L2 for overall face detection and localization.

Of note, the FaD confusion matrices contrast with the L2 VP ones, where there is a predisposition for each participant to respond with the same number of faces each time (either 0 or 1), producing the appearance of flat confusion matrices as opposed to diagonal lines. For the trials involving L2, a systematic bias was noted for participant S3, wherein they tended to respond "no faces" for almost all L2 trials. The confusion matrices for S1, S2, and S4 also adopt a mostly flat appearance, indicating that they defaulted to responding either zero or one faces. No participant responded that there were two faces present during two face trials, which is consistent with the way L2 VP method works and its inability to detect low-contrast objects. When two mannequins were present, one was always white and therefore not well-detected by L2 VP.

The settings of the NVPs could be adjusted for individual preferences regarding phosphene vision experience and additionally provide a sense of depth using brightness. Brighter phosphenes indicate a closer object compared with further objects, which elicit a fainter phosphene response. Moreover, each NVP could preferentially output a single phosphene, multiple phosphenes, or a bounding circle of phosphenes ("disc mode") for a target object. Although subjective preferences during the acclimatization phase varied for this cohort, a comparison of outlined shapes vs. solid shapes with Argus II recipients has previously reported some benefit to reducing the number of active phosphenes for a target shape (daCruz et al., IOVS, 2012, 53, ARVO E-Abstract, #5507; Luo et al., IOVS, 2014, 55, ARVO E-Abstract, #1834). The NVP methods were configured to 4-m detection ranges for all participants in this study, which is further than the distance of the target objects (1.5 m for face trials, 3.5 m for chairs trials), but within a common distance range for social interactions.

This study has demonstrated that NVP algorithms have improved detection, identification, and localization capabilities for specific real-world tasks in the laboratory environment. Therefore, a future multi-

modal VP approach, using a combination of NVP tools and the comprehensive VP L2, could potentially broaden the retinal implant user's ability to detect high- and low-contrast objects as well as provide some confidence with identifying certain task-specific objects.

Based on the findings discussed in this paper, FaD and ChD processing methods have been incorporated into the participants' existing bionic eye systems alongside other general purpose VP methods. Participants have been equipped and trained in the use of a new processing unit, which has a range of VP methods available. These VP can be selected as required via a nine-button keypad (3 × 3 configuration). Aside from the task-specific FaD and ChD NVP, the remaining buttons correspond with general purpose VP methods, including L2 and a range of new depth-based VP methods, with the final button reserved for muting the system. This enables participants control to switch between modes as needed in real-world environments.

In conclusion, the FaD and ChD NVP methods performed significantly better than L2 for the purpose of detecting, identifying, and localizing the presence of faces and available chairs in the laboratory setting. The results demonstrate that the functionality of an existing implanted suprachoroidal retinal prosthesis can be improved by updating the associated VP software. This means that retinal prosthesis recipients can experience improved social interactions in indoor settings. Future research will focus on integrating the NVP features with the comprehensive VP software in dynamic real-world outdoor settings.

Acknowledgments

Supported by NHMRC project grant (1082358), Australian Government Medical Research Future Fund BioMedTech Horizons Program Grant between MTPConnect MTP-IIGC Ltd to Bionic Vision Technologies Pty Ltd. Bionics Institute and the Centre for Eye Research Australia wish to acknowledge the support of the Victorian Government through its Operational Infrastructure Support Program and for generous support from the estate of the late Dr Brian Entwisle.

Disclosure: **L. Lombardi**, Bionic Vision Technologies (F); **M.A. Petoe**, Bionic Vision Technologies (F), Bionics Institute (P); **L. Moussallem**, Bionic Vision Technologies (F); **M. Kolic**, Bionic Vision Technologies (F); **E.K. Baglin**, Bionic Vision Technologies (F); **S. Stefopoulos**, Bionic Vision Technologies (E); **X. Battiwalla**, Bionic Vision Technologies (E); **J.G. Walker**, None; **N. Barnes**, Bionic Vision Technologies (F), Australian National University (P); **C.J. Abbott**, Bionic Vision Technologies (F); **P.J. Allen**, Bionic Vision Technologies (F), Centre for Eye Research Australia (P)

* CJA and PJA are co-senior authors of this work.

See the Appendix for the members of the Bionics Institute and Centre for Eye Research Australia Retinal Prosthesis Consortium.

References

1. Ail D, Malki H, Zin EA, Dalkara D. Adeno-associated virus (AAV) - based gene therapies for retinal diseases: where are we? *Appl Clin Genet*. 2023;16:111–130.
2. Botto C, Rucli M, Tekinsoy MD, Pulman J, Sahel JA, Dalkara D. Early and late stage gene therapy interventions for inherited retinal degenerations. *Prog Retin Eye Res*. 2022;86:100975.
3. Keseru M, Feucht M, Bornfeld N, et al. Acute electrical stimulation of the human retina with an epiretinal electrode array. *Acta Ophthalmol*. 2012;90(1):e1–e8.
4. Klauke S, Goertz M, Rein S, et al. Stimulation with a wireless intraocular epiretinal implant elicits visual percepts in blind humans. *Invest Ophthalmol Vis Sci*. 2011;52(1):449–455.
5. Luo YH, da Cruz L. The Argus((R)) II retinal prosthesis system. *Prog Retin Eye Res*. 2016;50:89–107.
6. Chow AY, Bittner AK, Pardue MT. The artificial silicon retina in retinitis pigmentosa patients (an American Ophthalmological Association thesis). *Trans Am Ophthalmol Soc*. 2010;108:120–154.
7. Lorach H, Goetz G, Mandel Y, et al. Performance of photovoltaic arrays in-vivo and characteristics of prosthetic vision in animals with retinal degeneration. *Vision Res*. 2015;111(Pt B):142–148.
8. Rizzo JF, 3rd. Update on retinal prosthetic research: the Boston Retinal Implant Project. *J Neuroophthalmol*. 2011;31(2):160–168.
9. Stingl K, Zrenner E. Electronic approaches to restitute vision in patients with neurodegenerative diseases of the retina. *Ophthalmic Res*. 2013;50(4):215–220.
10. Abbott CJ, Nayagam DAX, Luu CD, et al. Safety studies for a 44-channel suprachoroidal retinal prosthesis: a chronic passive study. *Invest Ophthalmol Vis Sci*. 2018;59(3):1410–1424.

11. Nayagam DA, Williams RA, Allen PJ, et al. Chronic electrical stimulation with a suprachoroidal retinal prosthesis: a preclinical safety and efficacy study. *PLoS One*. 2014;9(5):e97182.
12. Petoe MA, Titchener SA, Kolic M, et al. A second-generation (44-channel) suprachoroidal retinal prosthesis: interim clinical trial results. *Transl Vis Sci Technol*. 2021;10(10):12.
13. Fujikado T, Kamei M, Sakaguchi H, et al. One-year outcome of 49-channel suprachoroidal-transretinal stimulation prosthesis in patients with advanced retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2016;57(14):6147–6157.
14. Lee S, Ahn JH, Seo JM, Chung H, Cho DI. Electrical characterization of 3D Au microelectrodes for use in retinal prostheses. *Sensors (Basel)*. 2015;15(6):14345–14355.
15. Allen PJ. Retinal prostheses: where to from here? *Clin Exp Ophthalmol*. 2021;49(5):418–429.
16. Karapanos L, Abbott CJ, Ayton LN, et al. Functional vision in the real-world environment with a second-generation (44-channel) suprachoroidal retinal prosthesis. *Transl Vis Sci Technol*. 2021;10(10):7.
17. Petoe MA, Abbott CJ, Titchener SA, et al. A second-generation (44-channel) suprachoroidal retinal prosthesis: a single-arm clinical trial of feasibility. *Ophthalmol Sci*. 2025;5(1):100525.
18. Allen PJ, Kolic M, Baglin EK, et al. Second-generation (44-channel) suprachoroidal retinal prosthesis: surgical stability and safety during a 2-year clinical trial. *Clin Exp Ophthalmol*. 2025, https://doi.org/10.1111/ceo.14502.
19. Barnes N, Scott AF, Lieby P, et al. Vision function testing for a suprachoroidal retinal prosthesis: effects of image filtering. *J Neural Eng*. 2016;13(3):036013.
20. Abbott CJ, Baglin EK, Kolic M, et al. Interobserver agreement of electrode to retina distance measurements in a second-generation (44-channel) suprachoroidal retinal prosthesis. *Transl Vis Sci Technol*. 2022;11(9):4.
21. Titchener SA, Kvansakul J, Shivdasani MN, et al. Oculomotor responses to dynamic stimuli in a 44-channel suprachoroidal retinal prosthesis. *Transl Vis Sci Technol*. 2020;9(13):31.
22. Titchener SA, Nayagam DAX, Kvansakul J, et al. A second-generation (44-channel) suprachoroidal retinal prosthesis: long-term observation of the electrode-tissue interface. *Transl Vis Sci Technol*. 2022;11(6):12.
23. Saunders AL, Williams CE, Heriot W, et al. Development of a surgical procedure for implantation of a prototype suprachoroidal retinal prosthesis. *Clin Exp Ophthalmol*. 2014;42(7):665–674.
24. Petoe MA, Abbott CJ, Titchener SA, et al. A second-generation (44-channel) suprachoroidal retinal prosthesis: a single-arm clinical trial of feasibility. *Ophthalmol Sci*. 2024;100525:100525.
25. Mangiafico S. *Rcompanion: functions to support extension education program evaluation Version 2.5.0. Rutgers Cooperative Extension*. New Brunswick, NJ: Rutgers Cooperative Extension; 2025.
26. R Core Team. *R: a language and environment for statistical computing. Version 4.4.2. R Foundation of Statistical computing*. Vienna, Austria: The R Foundation for Statistical Computing; 2024.
27. Sadeghi R, Kartha A, Barry MP, et al. Glow in the dark: using a heat-sensitive camera for blind individuals with prosthetic vision. *Vision Res*. 2021;184:23–29.
28. Han N, Srivastava S, Xu A, Klein D, Beyeler M. Deep learning–based scene simplification for bionic vision. *Proceedings of the Augmented Humans International Conference 2021*. Rovaniemi, Finland: Association for Computing Machinery; 2021:45–54.
29. Sanchez-Garcia M, Martinez-Cantin R, Guerrero JJ. Semantic and structural image segmentation for prosthetic vision. *PLoS One*. 2020;15(1):e0227677.

Appendix. The Bionics Institute and Centre for Eye Research Australia Retinal Prosthesis Consortium

For this publication, the Bionics Institute and Centre for Eye Research Australia Retinal Prosthesis Consortium consists of (in alphabetical order): Lauren N. Ayton,[1,2,3] Peter J. Blamey,[4] Robert J. Briggs,[5,6] Owen Burns,[4] Stephanie B. Epp,[4] Dean Johnson,[5] Lewis Karapanos,[2,3] William G. Kentler,[8] Jessica Kvansakul,[4] Chi D. Luu,[2,3] Hugh J. McDermott,[4,9] Ceara McGowan,[4] Myra B. McGuinness,[2,10] Rodney E. Millard,[4] David A.X. Nayagam,[2,11] Peter M. Seligman,[4,9] Robert K. Shepherd,[4,9] Mohit N. Shivdasani,[4,12] Nicholas C. Sinclair,[4,9] Patrick C.Thien,[4,9] Samuel A. Titchener,[4,9] Joel Villalobos,[4,9] Chris E. Williams,[4,9,12] Jonathan Yeoh,[5,13] and Kiera A. Young.[2]

[1]Department of Optometry and Vision Sciences, University of Melbourne, Melbourne, Victoria, Australia

[2]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Victoria, Australia
[3]Department of Surgery (Ophthalmology), University of Melbourne, Melbourne, Victoria, Australia
[4]Bionics Institute, East Melbourne, Victoria, Australia
[5]Specialised Orientation and Mobility, Birregurra, Victoria, Australia
[6]Department of Otolaryngology, Royal Victorian Eye and Ear Hospital, East Melbourne, Victoria, Australia
[7]Otolaryngology, Department of Surgery, University of Melbourne, Melbourne, Victoria, Australia
[8]Department of Biomedical Engineering, University of Melbourne, Melbourne, Victoria, Australia
[9]Medical Bionics Department, University of Melbourne, Melbourne, Victoria, Australia
[10]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia
[11]Department of Clinical Pathology, University of Melbourne, Melbourne, Victoria, Australia
[12]Graduate School of Biomedical Engineering, University of New South Wales, New South Wales, Australia
[13]Vitreoretinal Unit, Royal Victorian Eye and Ear Hospital, East Melbourne, Victoria, Australia