

Performance of complex visual tasks using simulated prosthetic vision via augmented-reality glasses

Elton Ho

Jack Boffa

Daniel Palanker

Department of Physics, Stanford University,
Stanford, CA, USA

Hansen Experimental Physics Laboratory,
Stanford University, Stanford, CA, USA



Hansen Experimental Physics Laboratory,
Stanford University, Stanford, CA, USA

Hansen Experimental Physics Laboratory,
Stanford University, Stanford, CA, USA
Department of Ophthalmology, Stanford University,
Stanford, CA, USA

Photovoltaic subretinal prosthesis is designed for restoration of central vision in patients with age-related macular degeneration (AMD). We investigated the utility of prosthetic central vision for complex visual tasks using augmented-reality (AR) glasses simulating reduced acuity, contrast, and visual field. AR glasses with blocked central 20° of visual field included an integrated video camera and software which adjusts the image quality according to three user-defined parameters: resolution, corresponding to the equivalent pixel size of an implant; field of view, corresponding to the implant size; and number of grayscale levels. The real-time processed video was streamed on a screen in front of the right eye. Nineteen healthy participants were recruited to complete visual tasks including vision charts, sentence reading, and face recognition. With vision charts, letter acuity exceeded the pixel-sampling limit by 0.2 logMAR. Reading speed decreased with increasing pixel size and with reduced field of view (7°–12°). In the face recognition task (four-way forced choice, 5° angular size) participants identified faces at >75% accuracy, even with 100 μm pixels and only two grayscale levels. With 60 μm pixels and eight grayscale levels, the accuracy exceeded 97%. Subjects with simulated prosthetic vision performed slightly better than the sampling limit on the letter acuity tasks, and were highly accurate at recognizing faces, even with 100 μm/pixel resolution. These results indicate feasibility of reading and face recognition using prosthetic central vision even with 100 μm pixels, and performance improves further with smaller pixels.

Introduction

Age-related macular degeneration (AMD) is a leading cause of untreatable visual impairment. With the current prevalence of 8.7% worldwide, AMD is projected to affect almost 200 million people in 2020, and its prevalence is growing with the population aging (Friedman, Tomany, McCarty, & De Jong, 2004; Wong et al., 2014). Patients with advanced atrophic AMD (currently about 1% prevalence in Western countries (Friedman et al., 2004; Wong et al., 2014)) suffer from the loss of photoreceptors in the macula, leading to compromised central vision. Although high-resolution vision is lost, patients still can use their preserved peripheral vision and typically retain acuity no worse than 20/400. Therefore, restoration of central vision may be worthwhile only if the restored visual acuity exceeds the residual natural level.

In the healthy eye, photoreceptors convert incident light into electrical and chemical signals. The resultant neural signals are processed by the bipolar cells and other nonspiking neurons in the inner nuclear layer (INL) and advance to the retinal ganglion cells (RGC), which generate action potentials that propagate via optic nerve to the brain. Loss of photoreceptors in retinal degenerative diseases impairs the initial photo-transduction process, while the remaining retinal network remains intact, albeit with some rewiring (Humayun et al., 1999; Kim et al., 2002; Mazzoni, Novelli, & Strettoi, 2008).

Citation: Ho, E., Boffa, J., & Palanker, D. (2019). Performance of complex visual tasks using simulated prosthetic vision via augmented-reality glasses. *Journal of Vision*, 19(13):22, 1–10, <https://doi.org/10.1167/19.13.22>.

<https://doi.org/10.1167/19.13.22>

Received July 23, 2019; published November 26, 2019

ISSN 1534-7362 Copyright 2019 The Authors



Multiple approaches are being developed to address the loss of sight in retinal degeneration (Scholl et al., 2016), including gene therapy (Sengillo, Justus, Tsai, Cabral, & Tsang, 2016), cell transplantation (Lorach et al., 2019; Seiler et al., 2008), optogenetics (Barrett, Berlinguer-Palmini, & Degenaar, 2014), and electronic implants. In the latter case, an array of electrodes is placed at the stimulation site, such as the retina (D. Palanker & Goetz, 2018), optic nerve (Veraart, Wanet-Defalque, Gérard, Vanlierde, & Delbeke, 2003), lateral geniculate nucleus (LGN; Nguyen et al., 2016), or primary visual cortex (Lewis, Ackland, Lowery, & Rosenfeld, 2015). Electric current is injected into tissue to stimulate cells and thereby elicit visual perception. Upon electrode activation, patients report perceiving “bright spots,” termed phosphenes (Humayun et al., 2012; Stingl et al., 2015). The number of electrodes limits the amount of information deliverable, and electrode density restricts the highest possible resolution. In animal studies with photovoltaic retinal prosthesis, we demonstrated that grating acuity matches the pixel pitch with 55 (Ho, Lorach, Huang, et al., 2018) and 75 μm pixels (Lorach, Goetz, Smith, et al., 2015). Recent clinical trial of such implants (PRIMA, by Pixium Vision) having 100 μm pixels also demonstrated that prosthetic visual acuity in AMD patients is only 10%–30% below the sampling limit of 20/420 for the current pixel size (D. V. Palanker et al., 2019).

The PRIMA implant stimulates the first layer of neurons after photoreceptors (INL), and therefore elicited network-mediated retinal responses retain many features of the natural signal processing, including flicker fusion at high frequencies (>20 Hz; Lorach, Goetz, Mandel, et al., 2015; Lorach, Goetz, Smith, et al., 2015), adaptation to static images (Stingl et al., 2013), antagonistic center-surround organization of receptive fields with linear and nonlinear summation of its subunits (Ho et al., 2017). Patients with the implant can perceive lines as thin as the pixel pitch on the retina (i.e., single pixel in width) and identify letters with the minimum gap in the letter C of 1.1–1.3 pixels (D. V. Palanker et al., 2019).

Since AMD patients retain peripheral vision, they have little problem with ambulation. However, impaired reading and face recognition pose significant challenges in daily living (Mitchell & Bradley, 2006). To assess the spatial resolution, number of grayscale levels and the size of the implant required for these visual tasks, we simulated prosthetic central vision using augmented-reality glasses with a camera. Prosthetic vision was mimicked by controlled reduction in spatial resolution, contrast and visual field of the images projected on the built-in display. Here, we investigate how well healthy subjects can accomplish complex visual tasks, including reading and face recognition,

under various levels of image degradation. The results of this study predict the best possible clinical outcomes, as prosthetic vision in patients with retinal degeneration is likely worse than just pixelized natural vision at reduced contrast.

Psychophysics studies of simulated prosthetic vision were conducted in the past, but we find those results insufficient for predicting the outcomes with our current implant. With photovoltaic subretinal implant for restoration of central vision in AMD patients, simulation requires the following specifications: (a) pixel density >100 pixels/ mm^2 , (b) no gaps between phosphenes, (c) visual field in the range of 7° – 10° , and (d) eye scanning is allowed. Since previous studies did not address these specifications, we conducted a psychophysics study to assess the limits of visual performance of the PRIMA system and set the expectations for the upcoming clinical trials.

Methods

Subjects

Nineteen subjects (ages 18–74), all recruited from personnel at Stanford University, signed informed consent and participated in the current study. All subjects had self-reported normal vision, and their visual acuity was verified with both a Landolt C test and ETDRS chart prior to the experiments. For complex reading tasks, subjects were required to have native or near-native English proficiency. All subjects had limited or no prior experience with virtual or augmented-reality (AR) glasses. The study was approved by the Stanford IRB panel on human subject research and conducted according to the institutional guidelines, following the tenets of the Declaration of Helsinki.

Experimental setup

The experimental apparatus included two parts: a stimulus presentation system and AR glasses with the head-on display and an image processing unit (Figure 1).

The stimulus presentation system involved a 24" monitor (ASUS VS248H-P) controlled by a laptop computer (Thinkpad 25, Lenovo) using a PsychToolbox-based (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli & Vision, 1997) custom software in Matlab. This system was used to display stimulus and record subjects' responses (such as accuracy and time taken) via experimenter input. The monitor was placed 30" away from a chinrest, where the subjects would place

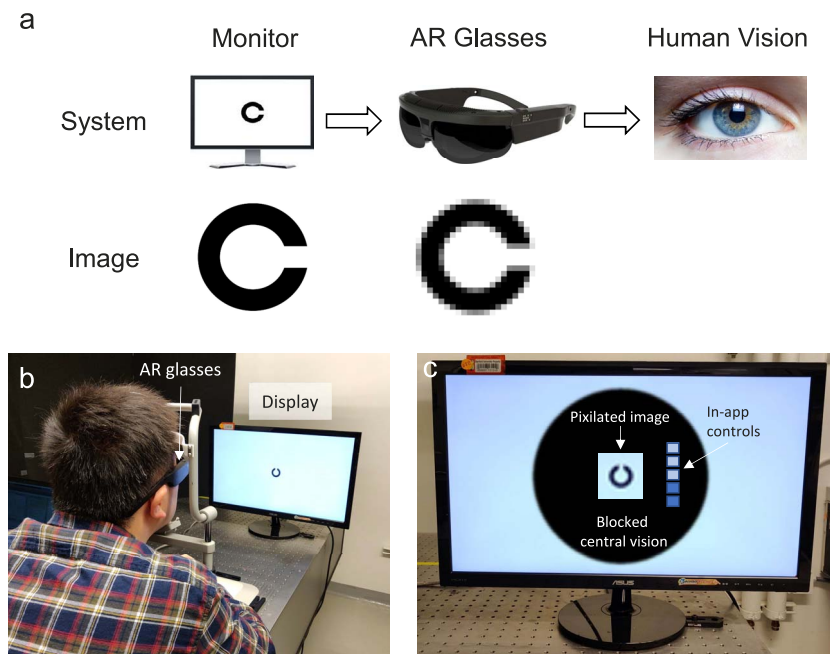


Figure 1. Experimental setup. (a) Schematic of the experimental setup. High resolution images are presented on a monitor. The front camera of the augmented-reality (AR) glasses captures the video stream. Custom software preloaded on the AR glasses adjusts the video quality to mimic prosthetic vision and displays it in the AR glasses. (b) A subject in front of the apparatus. (c) Illustration of vision through the AR glasses.

their head during an experiment. The monitor had resolution of 2400×1350 pixels, corresponding to 90.4 pixels per degree (ppd) of visual angle.

Camera (4 MP) mounted on the front of the AR glasses (ODG R-7, Osterhout Design Group, San Francisco, CA) captures a live video stream. Camera magnification was set to match the angular size of the natural vision. The data is then processed with an Android-based custom app in real time according to three user-defined parameters: pixelation (equivalent to 30–100 μm pixel size on the retina), number of grayscale levels (2–256), and field of view (FOV; 7° – 12°). The resultant video was presented on the display in the glasses (specs: 30° FOV, 720 p, 80 fps). The latency between the camera and the display was minimal due to fast video processing. In a typical AR display system, the integrated display is transparent, so that presented visual information can fuse with the passthrough background (hence, “augmented” reality).

To mimic vision loss in AMD patients, an area on the glasses corresponding 20° of central vision, was blocked with black opaque tape for both eyes. In this region, only the integrated display was visible, while outside that region, only natural peripheral vision was present (Figure 1c). Here we only assess monocular prosthetic vision, so the display was only switched on for the right eye, which incidentally corresponded to the dominant eye of all subjects. Visual information for small targets (Landolt C, letter identification, and face recognition) spanned a maximum 5° , so subjects used

electronic display exclusively, and could not benefit from moving their eyes outside the obscured area. Similarly, in the sentence reading tests with fonts of 1° or below, all the text was displayed behind the mask. As for sentences with larger font sizes, even though subjects could in theory have peeked outside, most of the sentence was still obscured. Subjects were instructed to vocalize the sentences sequentially, so looking at the few letters at the end of the sentence would not help with the task. Therefore, participants practically only read via the electronic display, which was confirmed through postinterviews.

The video processing was done with the OpenCV library, and the workflow was as following: A video frame was cropped to match the desired FOV. The frame was then converted to grayscale, downsized, and reupsized back to the original image size, resulting in a tightly packed pixelated, grayscale image. We used the default nearest-neighbor interpolation for image transformation in Android. The pixelation here matched the desired pixel size on the retina, e.g., 100 μm pixels subtend 0.35° on the human retina. The gray color value for each pixel was then rounded to the nearest $255/(n - 1)$, where n is the number of grayscale levels.

Subjects were instructed to wear the AR glasses and learned to adjust pixel size, grayscale levels, and FOV using in-app controls. To familiarize our subjects with simulated prosthetic vision, they were instructed to look around the laboratory freely for a few minutes,

Task	Equivalent pixel size on retina (μm)	Number of grayscale levels	FOV ($^\circ$)
Letter acuity	30, 60, 100	2, 8	7
Sentence reading	30, 60, 100	8	7, 12
Face recognition	60, 100	2, 4, 8	7

Table 1. Parameters of the image processing used for each experiment.

and were also presented pictures of common animals, plants, and foodstuffs.

Procedures

We conducted three different experiments: (a) letter visual acuity, (b) sentence reading, and (c) face recognition. Parameters for simulated prosthetic vision are summarized in Table 1. Subjects were instructed to fixate their central vision to the center of the AR screen but were allowed to move their eyes and head, if desired. In all experiments, subjects vocalized their responses, which were recorded and timed by the experimenter. Typically, a full set of experiments could last up to 90 minutes. If a subject got tired, a new session for remaining tasks was scheduled.

Letter acuity

Subjects ($n = 13$ for 30 and 60 μm pixels; $n = 19$ for natural vision and 100 μm pixels) were asked to identify the orientation of the Landolt C, presented one at a time. If the subject could identify at least four out of five orientations of the same size, we reduced the letter size by 0.1 LogMAR units and repeated. The same experiment was conducted also with ETDRS letters in Sloan font. The smallest feature of these characters was one fifth of the letter size. Subjects were first tested for their visual acuity with normal or corrected-to-normal vision without AR glasses, and then with simulated prosthetic vision. As a point for comparison, we also computed the sampling limit for each prosthetic pixel size by calculating its geometric-equivalent visual acuity.

Sentence reading

Subjects ($n = 9$ for simple sentences; $n = 10$ for complex sentences) were asked to read aloud displayed sentences as fast as possible, following standard MNREAD protocol (<http://legge.psych.umn.edu/mnread-set>). Text in Arial font was presented in three lines, with approximately 20 characters per line. A new

sentence with reduced font size (-0.1 LogMAR) was displayed upon successful utterance (≤ 2 mistaken words per sentence). The font size was measured as the visual angle between the top of the letter “k” and the bottom of the letter “p.” In between the sentences, a fixation cross was shown in the center of the screen for 2 s to recenter the subjects’ vision. Subjects were first tested with their normal/corrected binocular vision, and then with simulated prosthetic vision with varying pixel size and FOV. Eight grayscale levels were used to match the maximum expectations from the previously reported rodent studies (Ho, Lorach, Goetz, et al., 2018) and results with Alpha IMS implant (Stingl et al., 2015). The reading speed (in words per minute, or WPM) for each sentence was recorded in software. We evaluated reading performance on three key metrics: reading acuity (RA, smallest resolvable sentence), maximum reading speed (MRS), and critical print size (CPS, smallest font size at which 90% MRS is reached).

The texts used can be classified into simple and complex sentences. Simple sentences were either composed in-house according to the MNREAD protocol, or taken from the MNREAD iPad App ©2017 (<https://itunes.apple.com/us/app/mnread/id1196638274?ls=1&mt=8>). Complex sentences were selected from the Manually Annotated Sub-Corpus (MASC) from the Open American national Corpus (OANC) (<http://www.anc.org/>) with three criteria: number of characters between 55 and 70, average word length between 5.5 and 6.5, and sentence capable of being segmented into three lines of similar length. Generally, simple MNREAD sentences have stand-alone context and involve vocabulary at elementary school level in the US (e.g., “He looked up at his mother and told her he was really happy.”), while complex sentences may incur more context with advanced vocabulary (e.g., “Good housekeeping contributes to safety and reliable results.”). Only subjects with native or near-native English level were selected for the complex reading task. Results for simple and complex sentences were cross-compared using a two-sample t test.

Face recognition

Subjects ($n = 19$ for 100 μm ; $n = 17$ for 60 μm) were shown a reference adult face and required to select one out of four other faces that matched the identity of the reference as fast as possible (Figure 2a). The correctness and time taken for each selection were recorded. A set of 10 trials were performed for each parameter combination, which were presented in a pseudorandom order to minimize learning effects.

Images of nonoccluded adult heads were randomly selected from the Face Place database (<http://www>).

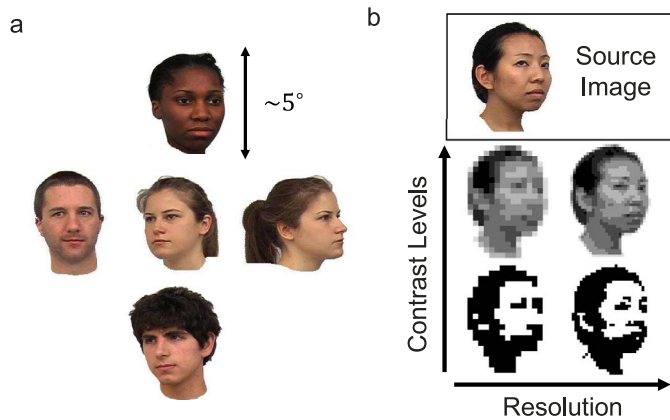


Figure 2. Face recognition task. (a) An example set of five faces presented. Subjects were asked to pick the face that matches the identity of the central person. Each face spanned approximately $5^\circ \times 5^\circ$. (b) Effects of the number of grayscale levels and resolution on an image.

tarrlab.org/). The database is licensed under a CC BY-NC-SA 3.0 Unported License. For the same identity, a set of images included different viewing angles and facial expressions, with the background cropped out. Generally, the most prominent features above the neck were visible, including hair style, skin tone, and both eyes. Images were resized and cropped to $5^\circ \times 5^\circ$. Five images were tiled as shown in Figure 2a, occupying a visual field of $16^\circ \times 16^\circ$. The reference image was placed at the center.

Results

Letter acuity

With both ETDRS letters and Landolt C, VA improved with reduced pixel size, as shown in Figure 3. The leftmost red data point indicates the subjects' normal or corrected visual acuity, whichever was better. VA measured by both testing paradigms agree with each other. The Landolt C test yielded slightly better VA than ETDRS letters by 0.05 logMAR, albeit insignificant (Supplement, Figure 1). Decreasing grayscale levels from 8 to 2 did not affect VA significantly. All measured VA were at least 0.2 logMAR better than the computed sampling limit for each pixel size. This could be attributed to oversampling by scanning and subjects looking for differences between undersampled letters. Letter recognition was also better than sampling limit and required only around 3 pixels per character width for all pixel sizes, agreeing with the 3–7 phosphenes per letter width reported by other studies (Dagnelie, Barnett, Humayun, & Thompson, 2006; Sommerhalder et al., 2003; Sommerhalder et al., 2004).

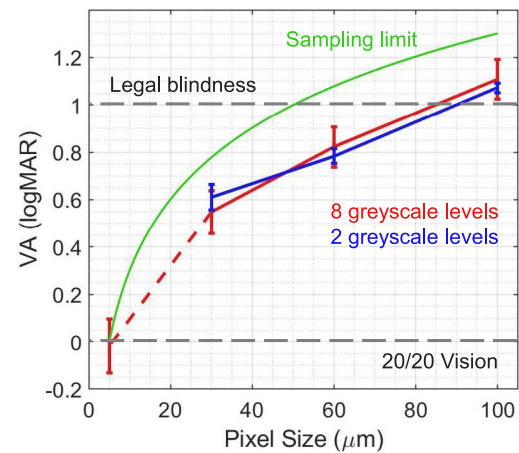


Figure 3. Letter acuity results ($n = 13$ for 30 and 60 μm pixels; $n = 19$ for natural vision and 100 μm pixels). The leftmost data point at 5 μm indicates visual acuity (VA) for natural vision of the subjects. Error bars are presented in terms of *SD*.

Most subjects self-reported that near the limit, they did not explicitly resolve the opening of a Landolt C. They employed a strategy where they scanned the object and identified the side of the blob that flickered more, through which correctly determining the orientation.

Sentence reading

With limited pixel size and FOV, reading speed with simulated prosthetic vision (Figure 4, green and red lines) was much slower than that with unobstructed natural vision (blue line). Reading acuity (RA) for both natural and prosthetic vision matched the corresponding letter acuity. As the font size increased above VA threshold, reading speed rapidly increased until the maximum reading speed (MRS) was reached at the critical print size (CPS). Further increase of the font size was detrimental, as fewer words and letters could fit in the FOV. For example, a nine-letter word of 1.5° font size (corresponding to 1.5° vertical height and 0.78° horizontal width allotted to each letter) can barely fit into 7° FOV. For all pixel sizes, CPS was around double the RA, and the smallest readable font size was about 2.5 pixels per letter width, slightly less than the letter acuity test and previous reports. The discrepancy can be attributed to the fact that in reading tasks, the loss in letter-by-letter information is compensated by contextual clues.

Generally, smaller pixels allowed for denser sampling, resulting in better RA, MRS, and CPS. Meanwhile, an increased FOV did not significantly affect RA, while raising reading speed with all font sizes greater than CPS ($t = 3.2$, $p = 0.005$ for MRS). The numerical results are summarized in Table 2.

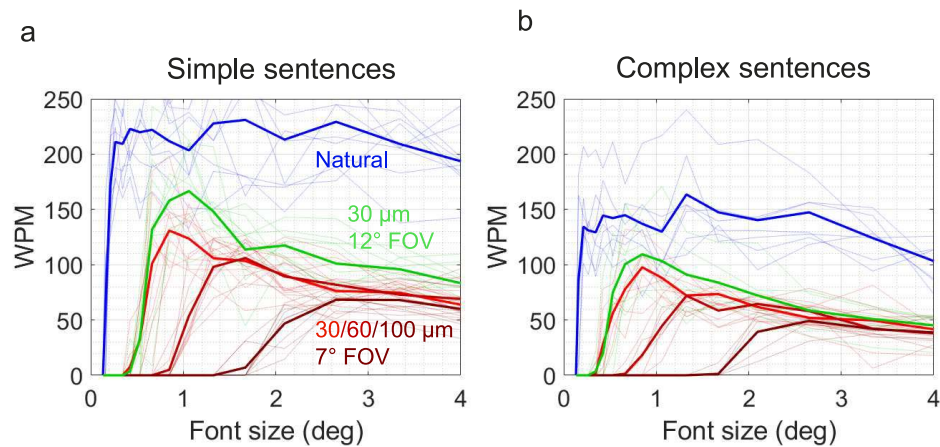


Figure 4. Sentence reading speed in words per minute (WPM). (a) Simple sentences. (b) Complex sentences. Faded lines represent individual measurements, and the bold lines represent the population mean.

General trends with complex sentences were the same, albeit at lower speed (Figure 4b and Table 3). However, the effect of FOV on MRS became insignificant (e.g., $t = 1.45$, $p = 0.156$ for $30 \mu\text{m}/12^\circ$). Counterintuitively, RA and CPS were slightly better (smaller) for complex sentences than for simple ones, possibly due to the word predictability in context-rich sentences.

Face recognition

For all pixel sizes and grayscale levels, subjects could achieve above 75% accuracy on average, significantly higher than random choice (25%; Figure 5). While faces were nearly instantaneously recognizable with natural vision, more than 5 seconds was needed with simulated vision, since scanning was required to observe all faces due to limited visual field. Increasing number of grayscale levels and reducing pixel size both improved accuracy and time taken for face recognition. In Figure 5c, response times are normalized to that for $100 \mu\text{m}$ pixels and eight grayscale levels. A decrease in pixel size from 100 to $60 \mu\text{m}$ shortened the response time by around 20% ($p < 0.025$ for all grayscale levels, two-

Parameters (pixels/FOV)	RA ($^\circ$)	MRS (WPM)	CPS ($^\circ$)
$100 \mu\text{m} / 7^\circ$	1.60 ± 0.15	58 ± 10	3.0 ± 0.6
$60 \mu\text{m} / 7^\circ$	0.81 ± 0.08	77 ± 16	1.7 ± 0.7
$30 \mu\text{m} / 7^\circ$	0.46 ± 0.10	106 ± 22	1.0 ± 0.6
$30 \mu\text{m} / 12^\circ$	0.45 ± 0.06	141 ± 27	1.1 ± 0.2
Natural	0.14 ± 0.01	201 ± 40	0.36 ± 0.07

Table 2. Reading acuity (RA), maximum reading speed (MRS), and critical print size (CPS) for reading MNREAD sentences using simulated prosthetic vision. Notes: All errors are reported as standard deviation.

sample t test). There was no significant difference in accuracy between 60 and $100 \mu\text{m}$ pixels.

Discussion

Letter acuity and reading speed are the most common metrics for assessment of the quality of vision, especially for low vision patients (Rubin, 2013). We added a face recognition task since it is of high priority for patients with atrophic AMD (Taylor, Hobby, Binns, & Crabb, 2016). Many psychophysics studies with simulated prosthetic vision were designed to investigate potential capabilities of implants with various numbers of pixels (Dagnelie et al., 2006; Hayes et al., 2003; Irons et al., 2017; Shannon, 1992). Recent clinical results with photovoltaic subretinal prosthesis having $100 \mu\text{m}$ pixels (PRIMA, Pixium Vision) confirmed that prosthetic acuity in AMD patients, measured using Landolt C test, nearly matches the pixel pitch (D. V. Palanker et al., 2019). Moreover, recent measurements with $55 \mu\text{m}$ pixels in rats demonstrated that grating acuity matches the pixel

Parameters (pixels/FOV)	RA ($^\circ$)	MRS (WPM)	CPS ($^\circ$)
$100 \mu\text{m} / 7^\circ$	1.63 ± 0.11	52 ± 7	2.7 ± 0.4
$60 \mu\text{m} / 7^\circ$	$0.71 \pm 0.11^*$	73 ± 10	1.7 ± 0.5
$30 \mu\text{m} / 7^\circ$	$0.36 \pm 0.05^*$	102 ± 18	0.80 ± 0.15
$30 \mu\text{m} / 12^\circ$	$0.37 \pm 0.09^*$	121 ± 33	$0.87 \pm 0.25^*$
Natural	0.14 ± 0.01	173 ± 34	0.32 ± 0.09

Table 3. Reading Acuity (RA), maximum reading speed (MRS), and critical print size (CPS) for reading complex sentences using pixelated vision. Notes: All errors are reported as standard deviation. Asterisk (*) indicates $p < 0.05$ (two-sample t test) compared to simple sentences with the same parameters.

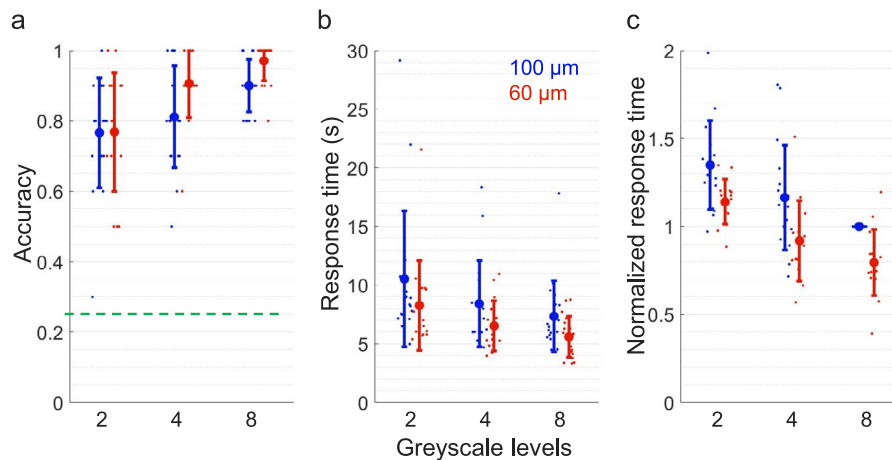


Figure 5. Face recognition. (a) Accuracy. (b) Response time. (c) Response time normalized to 100 μm pixels and eight grayscale levels. Each dot represents an independent measurement. Error bars are presented in terms of *SD*.

pitch of this size as well (Ho, Lorach, Huang, et al., 2018). Development of three-dimensional electrodes enables even smaller pixels, which might provide higher resolution in the future (Flores et al., 2018). To assess the minimum requirements of a system for restoration of central vision in AMD patients sufficient for reading and face recognition, we decided to evaluate its simulated performance as a function of three parameters: pixel size, field of view (FOV), and number of grayscale levels.

Previous studies with simulated vision used “phosphened” images (Chen, Hallum, Lovell, & Suaning, 2005; Dagnelie et al., 2006; Thompson, Barnett, Humayun, & Dagnelie, 2003). A dot with either a 2D-Gaussian or flat profile was displayed to simulate an activated pixel, while adjacent dots were spaced according to the pixel pitch of the implant, resulting in dark gaps between the simulated phosphenes (Chen, Suaning, Morley, & Lovell, 2009). However, in the PRIMA clinical study (D. V. Palanker et al., 2019), when viewing various line patterns, patients reported perceiving continuous lines, instead of a row of disconnected phosphenes. Therefore, in our study, we used tightly packed pixels, akin to those of a typical consumer monitor, with no dark gaps in between.

Another difference between the current study and previous ones is the choice of FOV. Since other implants were designed for inherited retinal degenerations which cause complete blindness, their functional FOV could be as large as 22° (Luo & da Cruz, 2016). However, geographic atrophy rarely exceeds 4 mm in diameter, and in order to avoid any damage to the adjacent healthy retina, the implant can cover only a part of the scotoma. Hence, subretinal implants for AMD are unlikely to exceed 3 mm in width, corresponding to approximately 10° of the visual angle. In the first feasibility study, the size of the PRIMA implant is 2 mm, corresponding to about 7° of the

visual field (D. V. Palanker et al., 2019). Therefore, we studied the effect of the FOV on reading speed in the range of 7° to 12° , while all the visual information for a letter acuity or face recognition tasks was packed within 5° of the visual angle.

When our subjects initially were unable to identify the orientation of small Landolt C, they were asked to guess without the experimenter affirming the answer. Typically, the subjects could correctly detect an extra line or two of the acuity chart, which explains their performance exceeding the sampling limit by about 0.2 LogMAR, as can be seen in Figure 3. This strategy is based on scanning the object and identifying a darker or a flickering size of the unresolved blob, which is sufficient for determining the Landolt C orientation. Such strategy can be used for other tasks within a small pool of target patterns, such as letter recognition, but is unlikely to help in identification of unknown objects and patterns.

It was repeatedly shown in the past that accuracy of the face recognition is highly dependent on image resolution, as summarized in (Irons et al., 2017). With 16×16 phosphenes per face over 9.4° visual field, and 10 levels of gray without scanning, subjects could differentiate faces with up to 84% accuracy (Chang, Kim, Shin, & Park, 2012), one of the highest reported. In another study with 24×24 phosphenes within 18° FOV, accuracy was 65%, and it reached 88% with 32×32 arrays (Wang et al., 2014). In the current study, focused on modeling small implants in the central macula, we used substantially smaller images (face spanning $5^\circ \times 5^\circ$) with higher pixel density, while the numbers of pixels per image were comparable to those in previous studies. We found that nearly perfect accuracy can be achieved at eight grayscale levels with 60 μm pixels, corresponding to a 24×24 grid. On top of using tightly packed pixels and allowing for head scanning, another likely explanation of improved

performance is that when the most prominent facial features lie within the fovea (<2 mm in diameter), subjects can spend less effort on scanning, and focus more on evaluating the facial details.

Interestingly, forced-choice face differentiation in our study required significantly fewer pixels than object recognition in a previous study (Jung, Aloni, Yitzhaky, & Peli, 2015). With $100\ \mu\text{m}$ pixels, corresponding to approximately 200 pixels per face, our subjects could differentiate faces at $>75\%$ accuracy. This is much less than about 560 pixels needed to recognize objects covering about 10° visual field on a de-cluttered background. The difference could be due to great simplification of the task when a reference is immediately available, compared to naming an object from a large pool of options. Another possibility is that faces could be a surprisingly easy class of images to discern. In a study involving different classes of objects and animals (Li, Hu, Chai, & Peng, 2012), subjects demonstrated $>80\%$ recognition rate on all images using 24×24 pixels. However, with 16×16 pixels, no one could recognize a car, but 90% could identify a dog, which coincides with the accuracy and parameters in our face recognition task. It is also important to keep in mind that in our study the faces we presented on a white background, while with a more cluttered natural background, two to three times more pixels maybe needed to achieve the same accuracy (Jung et al., 2015).

In conclusion, with simulated prosthetic vision in AR glasses, subjects demonstrated letter acuity slightly exceeding the sampling limit, and high efficacy in face recognition even with $100\ \mu\text{m}$ pixels. These results indicate that photovoltaic subretinal implants with $100\ \mu\text{m}$ pixels currently available for clinical testing may be helpful for reading and face recognition in patients who lost central vision due to retinal degeneration. As expected, smaller pixels significantly improve visual performance, and therefore, further reduction in pixel size may greatly enhance the outcomes in the future.

Keywords: prosthetic vision, augmented reality, visual acuity, reading, face recognition

Acknowledgments

This work was supported by the National Institutes of Health (Grants R01-EY-018608, R01-EY-027786), the Department of Defense (Grant W81XWH-15-1-0009), Stanford Institute of Neuroscience, and Research to Prevent Blindness.

Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, <http://www.tarrlab.org/>.

Funding provided by NSF award 0339122.

Commercial relationships: DP is consulting for Pixium Vision. DP's patents related to retinal prostheses are owned by Stanford University and licensed to Pixium Vision. EH and JB declare no competing financial interests.

Corresponding author: Elton Ho.

Email: eltonho@stanford.edu.

Address: Department of Physics, Stanford University, Stanford, CA, USA; Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA, USA.

References

- Barrett, J. M., Berlinguer-Palmini, R., & Degenaar, P. (2014). Optogenetic approaches to retinal prosthesis. *Visual Neuroscience*, *31*(4–5), 345–354.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Chang, M., Kim, H., Shin, J., & Park, K. (2012). Facial identification in very low-resolution images simulating prosthetic vision. *Journal of Engineering*, *9*(4), 046012.
- Chen, S. C., Hallum, L., Lovell, N., & Suaning, G. J. (2005). Visual acuity measurement of prosthetic vision: A virtual-reality simulation study. *Journal of Neural Engineering*, *2*(1), S135.
- Chen, S. C., Suaning, G. J., Morley, J. W., & Lovell, N. H. (2009). Simulating prosthetic vision: I. Visual models of phosphenes. *Vision Research*, *49*(12), 1493–1506.
- Dagnelie, G., Barnett, D., Humayun, M. S., & Thompson, R. W. (2006). Paragraph text reading using a pixelized prosthetic vision simulator: Parameter dependence and task learning in free-viewing conditions. *Investigative Ophthalmology & Visual Science*, *47*(3), 1241–1250.
- Flores, T., Huang, T. W., Lorach, H., Dalal, R., Lei, X., Kamins, T., . . . Palanker, D. V. (2018). Vertical walls surrounding pixels in subretinal space reduce stimulation threshold and improve contrast. *Investigative Ophthalmology & Visual Science*, *59*(9), 3975–3975.
- Friedman, D. S., Tomany, S. C., McCarty, C., & De Jong, P. (2004). Prevalence of age-related macular degeneration in the United States. *Arch ophthalmol*, *122*(4), 564–572.
- Hayes, J. S., Yin, V. T., Piyathaisere, D., Weiland, J. D., Humayun, M. S., & Dagnelie, G. (2003). Visually guided performance of simple tasks using

- simulated prosthetic vision. *Artificial Organs*, 27(11), 1016–1028.
- Ho, E., Lorach, H., Goetz, G., Laszlo, F., Lei, X., Kamins, T., . . . Palanker, D. (2018). Temporal structure in spiking patterns of ganglion cells defines perceptual thresholds in rodents with subretinal prosthesis. *Scientific Reports*, 8(1), 3145.
- Ho, E., Lorach, H., Huang, T. W., Lei, X., Flores, T., Kamins, T., . . . Palanker, D. V. (2018). Grating acuity of prosthetic vision in blind rats matches the pixel pitch of photovoltaic subretinal arrays below 50 μ m. *Investigative Ophthalmology & Visual Science*, 59(9), 3977–3977.
- Ho, E., Smith, R., Goetz, G., Lei, X., Galambos, L., Kamins, T. I., . . . Sher, A. (2017). Spatiotemporal characteristics of retinal response to network-mediated photovoltaic stimulation. *Journal of Neurophysiology*, 119(2), 389–400.
- Humayun, M. S., Dorn, J. D., da Cruz, L., Dagnelie, G., Sahel, J. A., Stanga, P. E., . . . Greenberg, R. J. (2012). Interim results from the international trial of Second Sight's visual prosthesis. *Ophthalmology*, 119(4), 779–788.
- Humayun, M. S., Prince, M., de Juan, E., Barron, Y., Moskowitz, M., Klock, I. B., & Milam, A. H. (1999). Morphometric analysis of the extramacular retina from postmortem eyes with retinitis pigmentosa. *Investigative Ophthalmology & Visual Science*, 40(1), 143–148.
- Irons, J. L., Gradden, T., Zhang, A., He, X., Barnes, N., Scott, A. F., & McKone, E. (2017). Face identity recognition in simulated prosthetic vision is poorer than previously reported and can be improved by caricaturing. *Vision Research*, 137, 61–79.
- Jung, J.-H., Aloni, D., Yitzhaky, Y., & Peli, E. (2015). Active confocal imaging for visual prostheses. *Vision Research*, 111, 182–196.
- Kim, S. Y., Sadda, S., Pearlman, J., Humayun, M. S., de Juan, E., Jr., Melia, B. M., & Green, W. R. (2002). Morphometric analysis of the macula in eyes with disciform age-related macular degeneration. *Retina*, 22(4), 471–477.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36(14), 1.
- Lewis, P. M., Ackland, H. M., Lowery, A. J., & Rosenfeld, J. V. (2015). Restoration of vision in blind individuals using bionic devices: A review with a focus on cortical visual prostheses. *Brain Research*, 1595, 51–73.
- Li, S., Hu, J., Chai, X., & Peng, Y. (2012). Image recognition with a limited number of pixels for visual prostheses design. *Artificial Organs*, 36(3), 266–274.
- Lorach, H., Goetz, G., Mandel, Y., Lei, X., Galambos, L., Kamins, T. I., . . . Palanker, D. (2015). Performance of photovoltaic arrays in-vivo and characteristics of prosthetic vision in animals with retinal degeneration. *Vision Research*, 111(Pt B), 142–148, <https://doi.org/10.1016/j.visres.2014.09.007>.
- Lorach, H., Goetz, G., Smith, R., Lei, X., Mandel, Y., Kamins, T., . . . Palanker, D. (2015). Photovoltaic restoration of sight with high visual acuity. *Nature Medicine*, 21(5), 476–482, <https://doi.org/10.1038/nm.3851>.
- Lorach, H., Kang, S., Bhuckory, M. B., Trouillet, A., Dalal, R., Marmor, M., & Palanker, D. (2019). Transplantation of mature photoreceptors in rodents with retinal degeneration. *Translational Vision Science & Technology*, 8(3):30.
- Luo, Y. H.-L., & da Cruz, L. (2016). The Argus® II retinal prosthesis system. *Progress in Retinal and Eye Research*, 50, 89–107.
- Mazzoni, F., Novelli, E., & Strettoi, E. (2008). Retinal ganglion cells survive and maintain normal dendritic morphology in a mouse model of inherited photoreceptor degeneration. *Journal of Neuroscience*, 28(52), 14282–14292.
- Mitchell, J., & Bradley, C. (2006). Quality of life in age-related macular degeneration: A review of the literature. *Health and Quality of Life Outcomes*, 4(1), 97.
- Nguyen, H. T., Tangutooru, S. M., Rountree, C. M., Kantzos, A. J., Tarlochan, F., Yoon, W. J., & Troy, J. B. (2016). Thalamic visual prosthesis. *IEEE Transactions on Biomedical Engineering*, 63(8), 1573–1580.
- Palanker, D., & Goetz, G. (2018). Restoring sight with retinal prostheses. *Physics Today*, 71(7), 26–32.
- Palanker, D. V., Le Mer, Y., Hornig, R., Buc, G., Deterre, M., Bismuth, V., & Sahel, J. A. (2019). Restoration of sight in geographic atrophy using a photovoltaic subretinal prosthesis. *Investigative Ophthalmology & Visual Science*, 60(9), 970–970.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Rubin, G. S. (2013). Measuring reading performance. *Vision Research*, 90, 43–51.
- Scholl, H. P., Strauss, R. W., Singh, M. S., Dalkara, D., Roska, B., Picaud, S., & Sahel, J.-A. (2016). Emerging therapies for inherited retinal degenera-

- tion. *Science Translational Medicine*, 8(368), 368rv6.
- Seiler, M. J., Thomas, B. B., Chen, Z., Wu, R., Sadda, S. R., & Aramant, R. B. (2008). Retinal transplants restore visual responses: trans-synaptic tracing from visually responsive sites labels transplant neurons. *European Journal of Neuroscience*, 28(1), 208–220.
- Sengillo, J. D., Justus, S., Tsai, Y. T., Cabral, T., & Tsang, S. H. (2016, December). Gene and cell-based therapies for inherited retinal disorders: An update. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* (Vol. 172, No. 4, pp. 349–366).
- 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). Simulation of artificial vision: I. Eccentric reading of isolated words, and perceptual learning. *Vision Research*, 43(3), 269–283.
- Sommerhalder, J., Rappaz, B., de Haller, R., Fornos, A. P., Safran, A. B., & Pelizzone, M. (2004). Simulation of artificial vision: II. Eccentric reading of full-page text and the learning of this task. *Vision Research*, 44(14), 1693–1706.
- Stingl, K., Bartz-Schmidt, K. U., Besch, D., Chee, C. K., Cottrill, C. L., Gekeler, F., . . . Zrenner, E. (2015). Subretinal visual implant alpha IMS—Clinical trial interim report. *Vision Research*, 111(Pt B), 149–160, <https://doi.org/10.1016/j.visres.2015.03.001>.
- Stingl, K., Bartz-Schmidt, K. U., Gekeler, F., Kusnyerik, A., Sachs, H., & Zrenner, E. (2013). Functional outcome in subretinal electronic implants depends on foveal eccentricity. *Investigative Ophthalmology & Visual Science*, 54(12), 7658–7665.
- Taylor, D. J., Hobby, A. E., Binns, A. M., & Crabb, D. P. (2016). How does age-related macular degeneration affect real-world visual ability and quality of life? A systematic review. *BMJ Open*, 6(12), e011504.
- Thompson, R. W., Barnett, G. D., Humayun, M. S., & Dagnelie, G. (2003). Facial recognition using simulated prosthetic pixelized vision. *Investigative Ophthalmology & Visual Science*, 44(11), 5035–5042.
- Veraart, C., Wanet-Defalque, M. C., Gérard, B., Vanlierde, A., & Delbeke, J. (2003). Pattern recognition with the optic nerve visual prosthesis. *Artificial Organs*, 27(11), 996–1004.
- Wang, J., Wu, X., Lu, Y., Wu, H., Kan, H., & Chai, X. (2014). Face recognition in simulated prosthetic vision: Face detection-based image processing strategies. *Journal of Neural Engineering*, 11(4), 046009.
- Wong, W. L., Su, X., Li, X., Cheung, C. M., Klein, R., Cheng, C. Y., & Wong, T. Y. (2014). Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: A systematic review and meta-analysis. *Lancet Global Health*, 2(2), e106–e116, [https://doi.org/10.1016/S2214-109X\(13\)70145-1](https://doi.org/10.1016/S2214-109X(13)70145-1).