Perceptual learning of prosthetic vision using video game training

Rebecca B. Esquenazi
Department of Psychology, University of Washington, Seattle, WA, USA

Kimberly Meier
College of Optometry, University of Houston, Houston, TX, USA

Michael Beyeler
Department of Computer Science and University of California, Santa Barbara, Santa Barbara, CA, USA
Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA, USA

Drake Wright
Department of Psychology, University of Washington, Seattle, WA, USA

Geoffrey M. Boynton
Department of Psychology, University of Washington, Seattle, WA, USA

Ione Fine
Department of Psychology, University of Washington, Seattle, WA, USA
School of Biomedical Sciences, Leeds, UK

A key limitation shared by both electronic and optogenetic sight recovery technologies is that they cause simultaneous rather than complementary firing within on- and off-center cells. Here, using "virtual patients"—sighted individuals viewing distorted input—we examine whether gamified training improves the ability to compensate for distortions in neuronal population coding. We measured perceptual learning using dichoptic input, filtered so that regions of the image that produced on-center responses in one eye produced off-center responses in the other eye. The Non-Gaming control group carried out an object discrimination task over five sessions using this filtered input. The Gaming group carried out an additional 25 hours of gamified training using a similarly filtered variant of the video game Fruit Ninja. Both groups showed improvements over time in the object discrimination task. However, there was no significant transfer of learning from the "Fruit Ninja" task to the object discrimination task. The lack of transfer of learning from video game training to object recognition suggests that gamification-based rehabilitation for sight recovery technologies may have limited utility and may be most effective when targeted on learning specific visual tasks.

Introduction

Outer retinal dystrophies—diseases that cause the death of photoreceptors—such as age-related macular degeneration (AMD) and retinitis pigmentosa (RP) are among the leading causes of blindness (Haim, 2002; Smith et al., 2001). Current treatments can slow disease progression but cannot restore photoreceptor function, and genetic interventions are likely to be limited to a subset of retinal dystrophies in the near future (Varela, de Guimaraes, Georgiou, & Michaelides, 2022). Other forms of treatment, such as retinal epithelial

Citation: Esquenazi, R. B., Meier, K., Beyeler, M., Wright, D., Boynton, G. M., & Fine, I. (2025). Perceptual learning of prosthetic vision using video game training. *Journal of Vision*, *25*(12):12, 1–14, https://doi.org/10.1167/jov.25.12.12.

https://doi.org/10.1167/jov.25.12.12 Received September 10, 2024; published October 7, 2025 ISSN 1534-7362 Copyright 2025 The Authors

(Kashani, 2022; Zarbin, Sugino, & Townes-Anderson, 2019), stem cell transplants (Shen, 2020), cell programming (Wang, Yang, Liu, & Qian, 2021), and a wide variety of other promising therapies (John, Quinn, Hu, Cehajic-Kapetanovic, & Xue, 2022), are in development. This article focuses on those sight restoration technologies designed to bypass normal photoreceptor function (Wood et al., 2019), including electronic retinal implants (Ayton et al., 2014; Ayton et al., 2020; da Cruz et al., 2016; Fujikado et al., 2016; Hornig et al., 2017; Lorach et al., 2015; Palanker, Le Mer, Mohand-Said, & Sahel, 2022; Rizzo et al., 2014; Saunders et al., 2014; Stingl et al., 2015), cortical implants (Beauchamp et al., 2020; Bosking et al., 2017; Fernandez et al., 2021; Troyk, 2017), and optogenetic technologies (AbbVie, 2022; Antolik, Sabatier, Galle, Fregnac, & Benosman, 2021; Korvasová et al., 2025; McClements, Staurenghi, MacLaren, & Cehajic-Kapetanovic, 2020; Sahel et al., 2021; Simunovic et al., 2019).

These sight recovery technologies are still grappling with a variety of significant technical and neurophysiological challenges. Retinal implants are currently the only devices approved for commercial implantation in patients, though a variety of cortical implants and optogenetic technologies are in clinical trials (Fernandez et al., 2021; Mohanty et al., 2025; Provansal, Marazova, Sahel, & Picaud, 2022; Sahel et al., 2021). Despite significant research efforts, to date no sight restoration technologies have proved capable of restoring useful vision in the majority of patients (Erickson-Davis & Korzybska, 2020). Many of these methods potentially elicit axonal stimulation that may distort phosphenes. Similarly, any form of technology involving external illumination must either track eye movements or deal with mismatches between the patients' head-mounted camera and eye-gaze position that result in a decreased ability to detect the correct location of stimuli (Caspi, Roy, Dorn, & Greenberg, 2017). Finally, electronic technologies require complex implantation surgery, optogenetic methods are likely to have difficulty achieving high transfection rates, and it is difficult to develop optogenetic molecules that have both fast kinetics and high light sensitivity.

One critical limiting factor, common to both electronic and optogenetic sight recovery technologies, is the inability to stimulate populations of cells in a naturalistic way. Current electronic, optogenetic, and optopharmacological (Van Gelder, 2015) approaches do not have the capacity to selectively stimulate cells with on-centered and off-centered receptive fields in a naturalistic complementary manner, in which activity in on-cells is accompanied by the suppression of off-cells and vice versa, as occurs in biologically natural vision.

Is it possible for plasticity to compensate for the distortions of the visual experience caused by unselective simultaneous stimulation of on- and off-cells? Previously, we tested the hypothesis that plasticity might compensate for abnormal population responses in the visual cortex (Esquenazi, Meier, Beyeler, Boynton, & Fine, 2021). With sighted participants, we presented dichoptic stimuli designed to elicit abnormal cortical population responses with a similar level of "scrambling," as would be expected from sight restoration technologies.

This was done by generating a spatial filter, *F*, using a radial checkerboard in Fourier space. Convolving an image *I* with this filter ($I * F$) filters out half of the spatial frequency and orientation information in the image, and convolving with the filter's complement, ($I * F'$), where $F' = I - F$, filters out the other half. The image $[I * F'] + [I' * F]$ was presented to one eye, while $[I * F] + [I' * F']$ was presented to the other, such that regions of the image that produced on-center responses in one eye produced off-center responses in the other eye and vice versa. Our participants exhibited significant performance improvements in an object discrimination task when trained with this distorted input over 14 training sessions. These results provided initial evidence that it may be possible for patients to adapt to the unnatural population responses produced by current electronic and optogenetic sight recovery technologies.

Here, we used the same dichoptic viewing manipulation to examine whether an engaging and dynamic training regimen—playing a video game—might result in an accelerated rate of generalizable learning that would transfer to a separate object recognition task.

In addition to this main experiment, we included a series of pre- and posttests focused on transfer of learning to stimulus variants of the object discrimination task, with the goal of elucidating the mechanisms underlying learning. Although these tests proved to be underpowered, the collective results were consistent with the hypothesis that the primary mechanism underlying improvements in performance in the object recognition task was learning to recognize "garbled" objects at a stage of processing after information from the two eyes has been combined.

Traditional perceptual learning paradigms use "low-level" visual features with naive observers repeatedly viewing and making judgments about stimuli that vary along a single dimension, such as contrast or spatial frequency. After the first few hundred trials (Karni & Sagi, 1993) learning tends to be slow, and improvements are typically restricted to the trained stimulus (Crist, Kapadia, Westheimer, & Gilbert, 1997; Fahle, 2005; Karni & Sagi, 1991) and task (Ahissar & Hochstein, 1997; Jeter, Dosher, Petrov, & Lu, 2009). These paradigms are inherently monotonous and time-consuming, requiring several hours to achieve even meager improvements in performance (Fine & Jacobs, 2002; Matthews, Liu, Geesaman, & Qian, 1999).

This has led to an increasing interest in "gamification" of perceptual learning. Action video games that feature fast-moving objects; a high degree of perceptual, cognitive, and motor load; increased engagement; unpredictable events; and significant peripheral processing (Green, Li, & Bavelier, 2010) seem to produce improvements in a variety of visual skills such as spatial resolution (Green & Bavelier, 2019), contrast sensitivity (Li, Polat, Makous, & Bavelier, 2009), visual search (Castel, Pratt, & Drummond, 2005), selective attention (Green & Bavelier, 2003), and object tracking (Trick, Jaspers-Fayer, & Sethi, 2005). The cortical mechanisms of improvement are thought to be twofold: First, action video games result in a highly engaged reward system through activation of dopaminergic pathways (Bavelier & Green, 2019), and second, action video games engage attentional pathways that aid in the selection of task-relevant information (Bavelier & Green, 2019). A good example of how video games are being used to target plasticity at relatively low levels of visual function is in the treatment of amblyopia in children (Fu, Wang, Li, Yu, & Yan, 2022; Gambacorta et al., 2018; Hussain, Astle, Webb, & McGraw, 2014).

One advantage of gamification as a training method for low vision (whether it be to adapt to vision loss or to regained prosthetic vision) is that it can be carried out relatively independently by the patient. Current methods of rehabilitation rely heavily on visual rehabilitation therapists and tend to focus on achieving specific goals, such as cooking safely or crossing the road. Effective gamification protocols would be more cost-effective and would reduce significant access barriers, since transportation is a difficulty for many blind individuals. One concern is that compliance with time-consuming home-based therapies is notoriously poor (Campbell et al., 2001; Essery, Geraghty, Kirby, & Yardley, 2017). However, there are reasons to be optimistic that an effective gamified visual rehabilitation protocol for blind individuals might benefit from higher levels of adherence than are typically observed. Blind individuals are often very motivated to regain function, especially in the context of the diminishment of independence and reduced access to activities that is often associated with blindness. It seems plausible that blind individuals will respond positively to the self-efficacy component of being able to independently engage in rehabilitation (Essery et al., 2017).

Our goal was to see whether video game training would substantially enhance the ability of participants to adapt to filtered visual input. Initial piloting found that a filtered version of a traditional action video game ("Mortal Kombat II," typical of those used for video game plasticity protocols) was impossibly difficult to play. We therefore created a custom video game based on the popular game Fruit Ninja and filtered it in real time using our dichoptic viewing manipulation. We chose Fruit Ninja because it contains many of the elements of more complex action video games. Participants are required to interact with fast-moving objects that appear unpredictably, focusing attention on identifying and tracking targets, while ignoring distractors.

After completion of a pretest, participants played our custom-coded and filtered video game for 25 hours. Every 5 hours of video game play, participants completed an object discrimination task, which served as a measure of learning transfer between video game training and the object discrimination task. A separate control group of participants did not play video games and instead completed the object discrimination task on roughly the same timeline as those in the video game group.

While there was significant learning for both the video game and the object discrimination task, there was little to no transfer of learning from the video game to the object discrimination task, suggesting that learning to interpret the unnatural on- and off-cell population responses likely to be produced by electronic and optogenetic sight recovery technologies may prove to be highly task specific.

Methods

This study was approved by the University of Washington's Institutional Review Board and carried out in accordance with the Code of Ethics of the Declaration of Helsinki. Informed consent was obtained before the start of the first experimental session.

Participants

Twenty naive observers (16 males) aged 19 to 32 years ($M = 25$) were recruited through word of mouth at the University of Washington. Participants were separated into two groups: a Non-Gaming control group ($n = 10$) and a Gaming ($n = 10$) group.

Binocular and monocular visual acuity was assessed using FrACT (Bach, 1996; Bach, 2007), and stereoacuity was assessed using the Randot Stereotest (Stereo Optical Co. Inc.). All observers had normal or corrected-to-normal visual acuity (defined as at least 0.2 logarithm of the minimum angle of resolution [logMAR] or 20/30 Snellen), interocular acuity differences no greater than 0.1 logMAR, and normal stereoacuity. The suppression check of the Randot Stereotest was used to confirm that no participant experienced abnormal suppression.

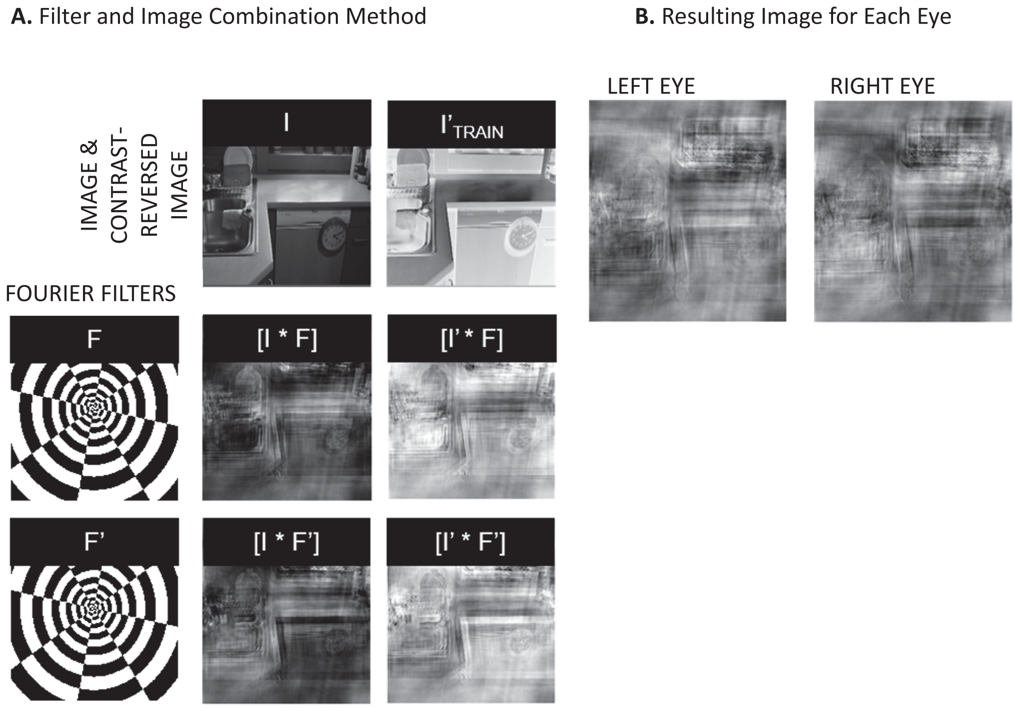

Figure 1. Example of filtering for dichoptic presentation, similar to Esquenazi et al. (2021). (**A**) The two upper panels show an example scene (I) and the contrast-reversed version of that scene (I′). The leftmost panels show two filters: F and F′. Filter images represent amplitudes in the Fourier domain, with spatial frequency increasing with distance from the center of the image and orientation changing with polar angle. The filters are paired complements, so the full spatial frequency and orientation content of the scenes is divided equally across the two filters. The lower middle panels show the convolution of the original (I) and contrast-reversed (I′) images with Fourier filters F and F′. (**B**) Examples of filtered images presented to left and right eyes for the training condition. Although these images do not resemble the perceptual experience of simultaneous on- and off-cell stimulation, interpretation of these images requires an analogous process of interpreting a garbled population response.

Equipment

Stimuli were presented dichoptically, using methods previously described in Esquenazi et al. (2021). A custom-built stereoscope consisted of two cold mirrors mounted on posts, rotated at a 45-degree angle to capture input from two separate 32-inch LED monitors and reflect it separately into each eye. Each monitor spanned 28.9 degrees, and these monitors were the only light source in the room. Monitors had a mean luminance of 80 cd/m^2. The stimuli had a mean luminance of 132 cd/m^2. Each stimulus spanned 768 × 768 pixels (8.84 degrees). Participants completed a nonius task at the beginning of each session to align the screens and account for any fixation disparities.

Filtering in real time was done using two computers. The first computer ran the video game. A KONA 4-Channel HDMI Capture Card was used to stream the HDMI output of the gaming computer to the stimulus processing computer at 30 Hz. Onboard the stimulus computer, the capture card directly passed each frame to a high-powered NVIDIA Quadro RTX 6000 graphics processing unit (GPU) without the involvement of the central processing unit, preserving the 30 Hz stream rate with minimal lag (∼1 frame). The GPU filtered each frame (as described below) using the Compute Unified Device Architecture (CUDA), a parallel programming framework developed by NVIDIA.

Filtering

Binarized radial checkerboard filters in Fourier space were used to present separate spatial frequency and orientation information to each eye (Figure 1).

As described previously (Esquenazi et al., 2021), the spatial frequency content of the filters, F_f, was defined as $F_{sf} = \sin(2\pi \mathrm{n} f_0 . f^{\frac{1}{n}})$, where f is the spatial frequency of the Fourier image, $f_0 = 13$ and controls the overall frequency of the radial rings, and n describes the increase in ring width as a function of spatial frequency. The orientation content of the filter, F_ϑ, was defined as $F_\vartheta = \alpha_0.\alpha$, where α is the orientation of the Fourier image, and α_0 defines the number of radial spokes.

We created two complementary final filters as the products of F_f and F_ϑ: $F = Ff \times F_\vartheta$, and $F' = -Ff \times F_\vartheta$. Finally, these filters were scaled and binarized to values 0 and 1. Since each filter was the complement of the other, the full spatial frequency and orientation content of both the original and the contrast-reversed scene were divided equally across the two filters and thus the two eyes.

The two upper-left panels of Figure 1 show an example scene, I, and the contrast-reversed version of that scene, $I' = 1 - I$. The leftmost panels show the two radial checkerboard Fourier filters F and F′. The original, I, and the contrast-reversed scene I′ were each converted into the Fourier domain, multiplied with one of the two Fourier filters, and then converted back to image space using the inverse Fourier transform. The top panels show original (left) and contrast-reversed (middle) images (right). The bottom two panels show the four examples of possible filtering: $I * F$, $I * F'$, $I' * F$, and $I' * F'$ (where * denotes two-dimensional convolution) for the object discrimination task stimuli.

Our radial checkerboard filter was built with sharp edges in the Fourier domain, which leads to image artifacts—often seen as "ringing" in the spatial domain. Our choice of a binary filter (rather than a filter with smooth edges in the Fourier domain) was motivated by the desire to minimize shared spatial frequency and orientation information within the images presented to each eye. These Fourier artifacts were relatively subtle, with a root mean square (RMS) contrast of ∼1/3 that of the original images, and pilot data suggested these artifacts were unlikely to be the primary cause of masking (Esquenazi et al., 2021).

Strong contours in our images can result in "striping" as a result of missing alternating frequency bands. The striping occurs at complementary frequencies in $[I * F]$ and $[I * F']$. The orientation, frequency, and strength of the striping depend on the orientation of the strong contours of the image in relationship to the filter bands. It is unlikely that participants learned to make use of these "striping" Fourier artifacts to perform the task because the object images in the task were always presented at random locations, orientations, and sizes, and the overall scaling of the background varied over each trial.

In the training paradigm, we presented the left eye with the sum of two filtered images, $[I * F'] + [I' * F]$, such that half the spatial frequency and orientation content was based on the original image, and the other half was based on the contrast-reversed image. In the right eye, we presented the sum of the complementary filtered images, $[I * F] + [I' * F']$.

Note that the sum, $[I * F] + [I * F']$, equals the original image I, and $[I' * F] + [I' * F']$ equals the original contrast-reversed image I′; thus, all the spatial frequency and orientation information of both the original and contrast-reversed images is preserved. Thus, with optimal decoding, stimuli are "lossless." The sum of the distorted images in each eye results in a blank image.

However, as described previously (Esquenazi et al., 2021), the normal pattern of population responses (wherein on- and off-cells with similar spatial frequencies and orientations tend to be highly correlated in their firing) is disrupted. In the absence of suppression, on- and off-cells with identical tuning profiles would be simultaneously stimulated by the combination of input from the left and the right eye, as occurs during electrical stimulation.

Object stimuli

Each phase of the study (pretest, video game, test for learning transfer, and posttest) consisted of a different set of 45 objects, for a total of 175 unique objects.

Learning protocol

As shown in Figure 2, our experiment contained three phases: (1) three sessions of the pretest object discrimination tasks (see Appendix 1 for a detailed description of these tasks), (2) five repeats of 5× video game training sessions (Gaming group only) + 1× object discrimination task session (all participants), and (3) three sessions of the posttest object discrimination tasks.

Participants in the Gaming group completed 25 video game sessions, each lasting 1 hour. During gaming sessions, participants played the filtered video game for at least 45 minutes, with 15 minutes allotted for a break. Each video game session was conducted on a separate day, with an average of three sessions per week. Every five sessions of gaming, participants completed one session of the object discrimination task, using identical filtering as used during gameplay.

Participants in the Non-Gaming group did not play video games and instead completed five sessions of the object discrimination task. The Non-Gaming group completed the object discrimination task every 7 to 10 days, so that each experimental group completed the object discrimination task sessions over a similar time period.

Because our goal at this stage was simply to see if adding gamification training to the object discrimination task enhanced learning, we did not have the control group perform an analogous "placebo" task. Had we found transfer of learning from the video game to the object discrimination task, a placebo task would have been necessary to demonstrate that this transfer of learning was specifically due to training with filtered input.

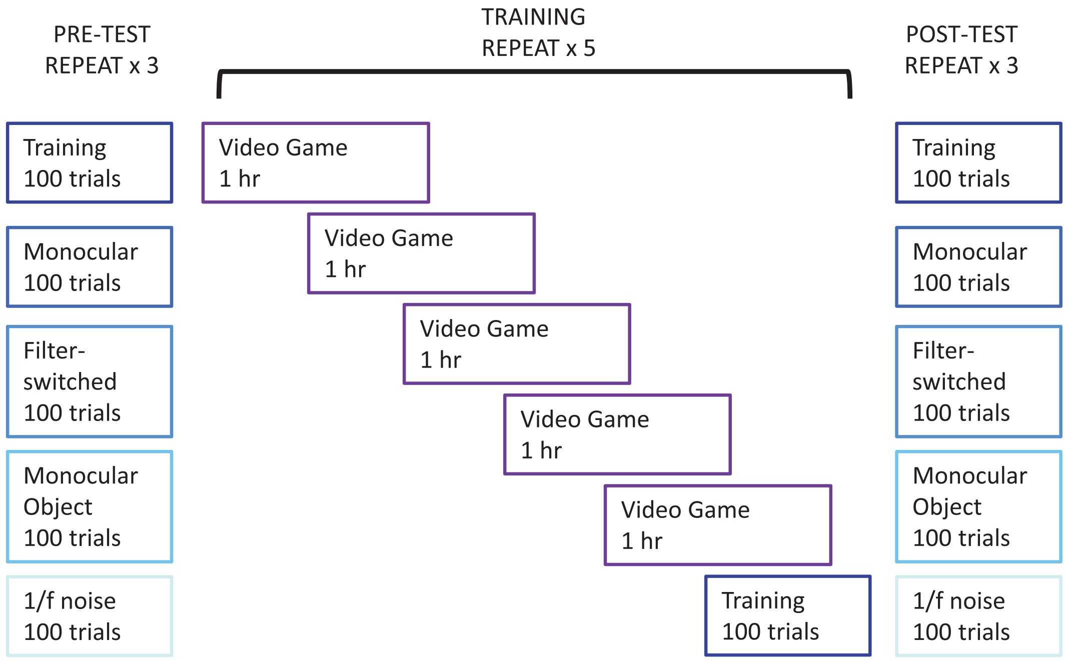

Figure 2. Schematic of learning and testing protocol. The object recognition task is shown in blue, and the video gaming task (Gaming group only) is shown in purple.

Video game training

Participants in the Gaming group played a custom-coded game using Godot open-source software (Linietsky, 2014) (Figure 3). The game was modeled after the well-known video game Fruit Ninja, in which fruits (apple, banana, etc.) are hurled onto the screen at various speeds. Players slice the fruits as they appear on screen with a blade controlled via touchscreen, while avoiding slicing distracting bombs that would end the game.

The goal of our game was similar to Fruit Ninja: identify Target Objects launched on screen, while avoiding Distractor Objects. In any given round,

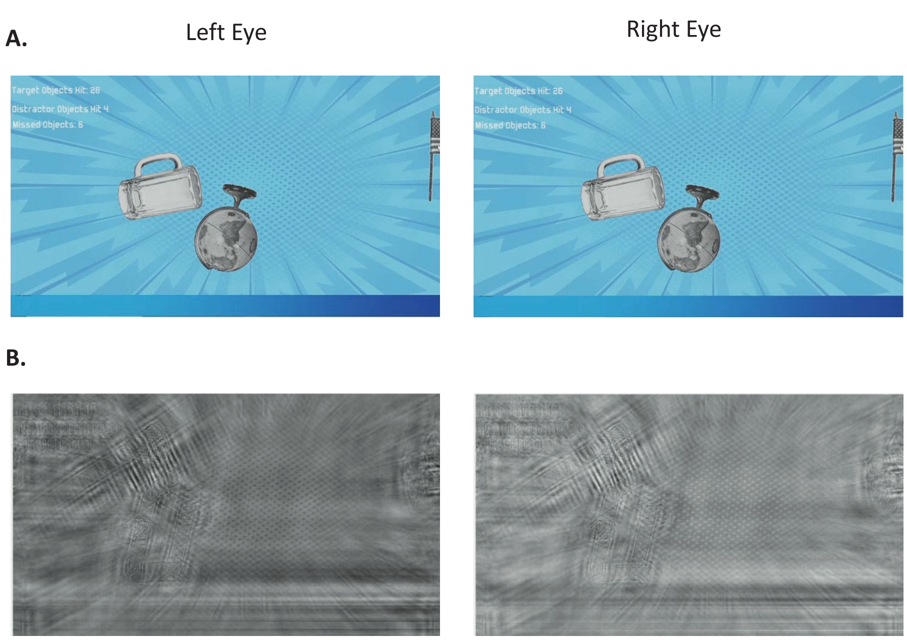

Figure 3. The Fruit Ninja training game, original (**A**) and filtered (**B**). Supplemental Materials include example movies for Panel A (unfiltered) and Panel B (filtered).

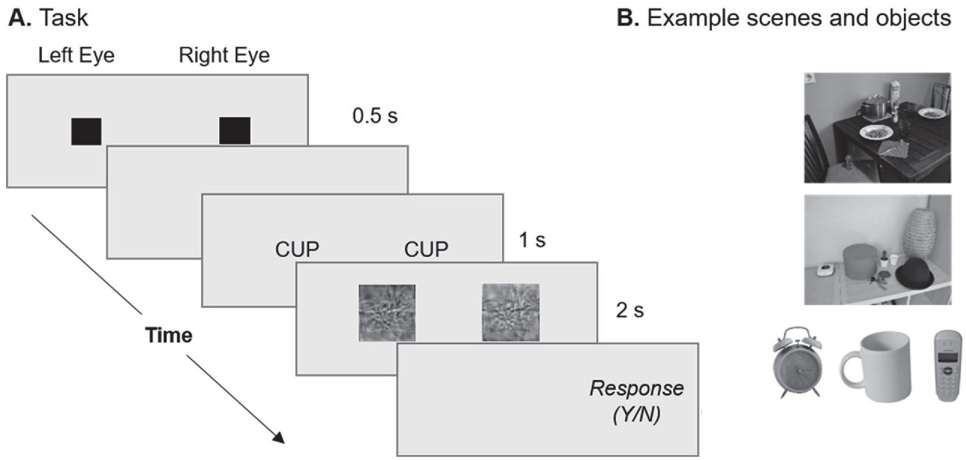

Figure 4. The object discrimination task, replicated from Esquenazi et al. (2021). (**A**) On each trial, the participant reported whether or not the cued object was present within the scene. (**B**) Examples of unfiltered scenes and objects from the SCEGRAM database.

a unique set of 6 (out of 45 possible objects) was randomly selected. Each object was launched into the scene from the bottom of the screen at a random location, and the object orientation was rotated by up to ±30 degrees.

Players always started each session with three lives at Level 1. At the beginning of each round, participants received directions to slash as many objects as possible (there were five Target Objects on each game round at Level 1) while avoiding slashing a single Distractor Object. While participants did not know which Target Objects would appear, they were presented with an unfiltered version of the Distractor Object to avoid prior to the start of each game round.

Objects were destroyed by clicking and dragging the computer mouse through the body of the object to create a slashing motion. If the player slashed a Target Object, they received positive auditory feedback and gained a "Target Object Point." If the player slashed a Distractor Object, they received negative auditory feedback and gained a "Distractor Object Point." Missing three Target Objects also resulted in negative auditory feedback and a "Distractor Object Point."

Players automatically advanced to the next level if they received 25 Target Object Points before dying. Receiving five Distractor Object Points caused the level to restart, and the player lost a life. When three lives were lost, the game automatically restarted at Level 1, and players received a "*Game Over*" notification.

The difficulty of the game varied within and between levels. As any given round progressed, objects were launched onto the screen with increasing velocity. Object sizes also varied depending on the gamer's level: As the player advanced over levels, the size of the objects decreased logarithmically for a range of sizes between ~2 and 6 degrees. Once the player reached Level 4, they were given two Distractor Objects to avoid and thus had just four Target Objects to slash. In total, there were six levels of difficulty. If players beat the game in any given training session, they were instructed to restart the game at Level 1.

Object discrimination task

Transfer of learning to the object discrimination task was assessed every five video game sessions. The object discrimination task was based on previous work (Esquenazi et al., 2021) (Figure 4). A brief fixation cue (0.5 seconds) began each trial. After a 0.5-second pause, a word cue told the participants what the target object was (e.g., "cup," "clock"). Following the word cue, a scene with an overlaid object (using a different set from the video game) was displayed for up to 2 seconds or until the participant responded with a key press.

Backgrounds consisted of 17 scenes of different household settings (e.g., kitchen, living room, bathroom, bedroom) from the SCEGRAM Database (Ohlschlager & Vo, 2017). To minimize image-specific learning, objects could take on one of six possible logarithmically spaced sizes ranging from 22.2% (1.96 degrees) to 66.7% (5.90 degrees) of the original object (768 × 768 pixels or 8.84 degrees). Each object was randomly located within the scene and was rotated by up to 30 degrees in either direction. Thus, there were over 13,000 unique images in each object-scene set.

Participants performed a two-alternative forced-choice (2AFC) object discrimination task, judging whether the scene contained the cued object. In each trial, there was a 50% chance that the scene contained the prompted object or a different distractor object. Auditory feedback was provided after each trial to indicate whether the answer was correct or incorrect.

Participants were not given specific instructions on where to look within the scene.

Tests of transfer of learning in pre- and posttests

During pre- and posttesting periods, we tested five different variants of the object discrimination stimulus (Appendix 1) to examine transfer of learning. These conditions were interleaved, in blocks of 10 trials. There were three pre- and posttest sessions, each containing 100 trials per condition, for a total of 300 trials per condition in both pre- and posttests for each condition.

Results

Video game training

Performance improvements during gaming were measured by calculating the average number of times per level that a participant was forced to restart the game ("*Game Over*") because they died three times. Figure 5 shows the average *Game Over* rate per level averaged over all participants, in each session. An average *Game Over* rate of zero would indicate that participants had entirely mastered that level. Participants mastered Levels 1 to 3 fairly quickly. By Session 11, the average *Game Over* rate for Level 4 was consistently less than 0.25. While training did improve performance for Levels 5 and 6, average *Game Over* rates remained high (ML5 = 0.61; ML6 = 0.94). The *Game Over* rate for Level 6 consistently remained above 0.9 across all participants, suggesting that participants never reached ceiling performance for that level.

Object discrimination task

Hits, misses, correct rejections, and false alarms from conditions within each session were converted into d-prime (d′) units (Green & Swets, 1966). Figure 6 shows d′ (Panel A) and normalized (Panel B) d′ scores for participants in the Non-Gaming and Gaming groups, over the five object discrimination sessions.

Two-way linear mixed effects regression analyses, using the lme4 package in R (R Core Team, 2022), were used to examine improvements in performance (in d′) across the five sessions of the object discrimination task for both the Non-Gaming and Gaming groups. Participants were treated as a random factor, and the session number was treated as a fixed factor.

Participant intercepts, d′ for Session 1, ranged from 0.51 to 2.59 ($M = 1.42$, $SD = 0.51$). We noticed a marginally significant difference in d′ for Session 1 between Non-Gaming ($M = 1.37$, $SD = 0.47$) and Gaming ($M = 1.83$, $SD = 0.59$) groups (Welch two-sample *t*-test, $t(17.15) = 1.90$, $p = 0.0750$, Cohen's d: 0.60). At this point, there was no experimental difference between the two groups. As described further in the Discussion, this difference likely reflects a differing level of commitment between the two experimental groups, since the Gaming group had

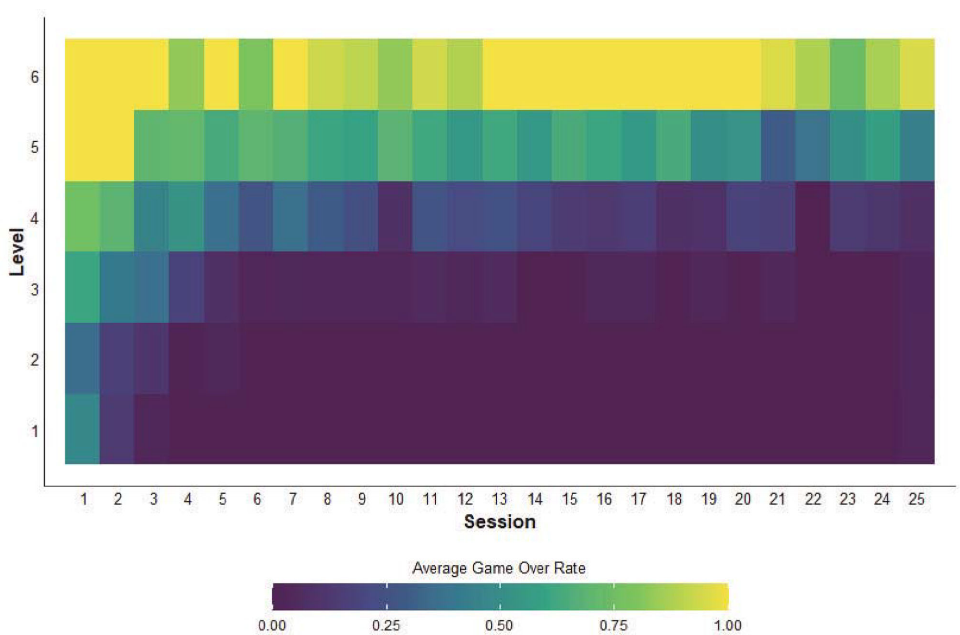

Figure 5. The average "*Game Over*" rate (i.e., the number of times gamers lost three lives and had to restart the game), for participants in each level, across all 25 gaming sessions. Data from individuals are shown in Appendix 2.

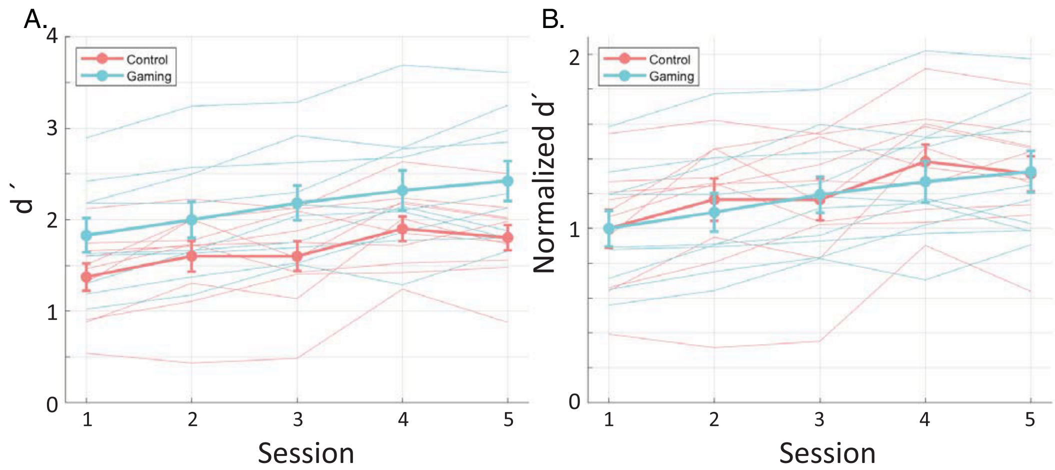

Figure 6. (**A**) d′ and (**B**) normalized d′ scores for participants in both the Non-Gaming and Gaming groups across the five object discrimination sessions (normalized d′=d′/d′session 1). Thick colored lines represent group means, while thin lines represent individual participants. Error bars reflect the standard error of the mean.

a higher proportion of graduate students than the Non-Gaming group. We compared d-prime values from the first object discrimination session between graduate students ($M = 1.92$, $SD = 0.56$) versus undergraduates ($M = 1.28$, $SD = 0.38$) and found a significant difference (Welch's t-test: $t(15.77) = 3.01$, $p = 0.0084$, Cohen's d: 0.95).

We assessed learning using a linear mixed-model analysis with time (the five sessions) and group (Non-Gaming vs. Gaming) as fixed factors, and participant and student status (graduate vs. undergraduate) as random-effect factors. We found a strong main effect for session ($F(1, 78) = 95.69$, $p < 0.0001$) but no significant effect for group ($F(1, 18) = 0.56$, $p = 0.2030$), and no significant group-by-session interaction ($F(1, 18) = 1.65$, $p = 0.2030$).

Thus, after accounting for the graduate student status, there was no significant difference in either the overall performance or in the learning rate between the Non-Gaming and Gaming groups. Participant d′ at Session 1 ranged from 0.51 to 2.59 with a mean value $M = 1.42$, $SD = 0.51$. The slope over sessions, an indicator of learning rate, was 0.1338 units of d′ per session across both groups. Over the course of five sessions, the predicted d′ increased from a mean of 1.64 to 2.17, an increase of 32.7%. For the Non-Gaming group, the slope was 0.1162 units of d′ per session, resulting in a 31.5% increase in d′ between the first and last sessions, and for the Gaming group, the slope was 0.1513, resulting in a 33.0% increase in performance.

A power analysis on the F-test group by object discrimination task session interaction was conducted using R. With 10 participants/group, a significance criterion of $\alpha = 0.05$, and power = 0.80, we would expect adequate power to find a significant group-by-session interaction with an effect size of $\eta^2 = 0.2408$. Given the variance in our data set, this is the equivalent of a slope of 0.1162 in the Non-Gaming group and a slope of 0.1928 in the Gaming group. Since there were five video gaming sessions for every object discrimination session, this would correspond to an hour of video game training session resulting in a 0.0153 generalizable improvement in d′ (less than 0.5% correct). We felt that effect sizes smaller than this would be of little clinical utility.

Finally, we also examined transfer of leaning to different variants of the object discrimination stimulus, as described in Appendix 1. Although underpowered, collectively these tests found complete transfer of learning across a variety of stimulus distortions, suggesting that improvements in performance were not contingent on the specifics of how the images were filtered.

Discussion

While our dichoptic images did not simultaneously stimulate on- and off-pathways per se, participants viewed input that likely caused a similar level of disruption to early population responses, as described previously (Esquenazi et al., 2021). In this previous study, we showed that it is possible to learn to discriminate everyday objects with prolonged training using this distorted input.

Here we asked whether playing an interactive video game with the distorted input might provide an enhanced learning experience. Specifically, we asked whether gaming might lead to generalizable learning that transfers to a novel set of objects.

Effectiveness of video game training

While we expected Non-Gaming participants to display some improvement in the object discrimination task due to repeated stimulus exposure, we predicted that the additional exposure to the filtered input during video game training might result in Gamers outperforming Non-Gamers on the object discrimination task. While we saw clear learning effects for both the video gaming and the object discrimination task, there was little or no transfer of learning between the two tasks. After accounting for graduate student status, we saw little evidence of differences in either overall performance or learning between Non-Gaming and Gaming groups. The negligible transfer of learning between the two tasks suggests that learning to compensate for abnormal early visual population responses may be highly task specific.

As described above, tasks that train on low-level stimulus features such as orientation and spatial frequency (Fiorentini & Berardi, 1980) typically lead to highly specific learning that fails to transfer across eye of origin or to novel stimuli, tasks, or retinal locations (Fahle, 2005; Karni & Sagi, 1991). Similarly, previous work (Ni & Maunsell, 2010) has shown that macaques slowly improve in their ability to detect electrical stimulation—and that this improvement interfered with the ability to detect light stimuli—suggesting that improved "decoding" of unnatural stimulation with training is possible. However, this learning was carried out with a simple detection task, was relatively slow, and was highly specific to the trained location.

In contrast, training using action video games can show a remarkably high rate of transfer across a wide range of visual tasks (Bavelier & Green, 2019; Dye, Green, & Bavelier, 2009; Green et al., 2010; Li et al., 2009). While mechanisms of generalizable learning as a result of video game training are still under investigation, it has been proposed that the high transfer rate to standard psychophysical tasks might be attributed to shared visual demands between the transfer task and the trained video game (Oei & Patterson, 2015).

Our "Fruit Ninja" video game was designed to maximize generalizable learning. The task required identifying object images that varied in size, location, and rotation, while progressively more difficult—conditions that all foster generalizable learning (Green et al., 2010). Further, participants appeared to master the game from Levels 1 to 4 by at least Session 15, indicating substantial learning during training (see Figure 5). Despite this, we saw negligible transfer from the Fruit Ninja to the object discrimination task.

This is quite surprising given that both tasks involved discriminating between common household objects. Despite this similarity between the two tasks, it seems that our training and testing tasks did not contain enough common elements to produce transfer of learning between them. One difference between the two tasks is the relationship between the objects and the background. Our video game required identification of objects quickly and simultaneously rotating in motion over a relatively plain static background. In contrast, the objects in our discrimination task were stationary (although the entire image did pan across the screen on each trial) with respect to cluttered backgrounds that varied across each trial.

A second possible reason why we may have failed to see transfer of learning is that, despite our best efforts (including reporting performance and providing catchy game music), playing a visually degraded version of Fruit Ninja in a dark room through a stereoscope is far less compelling than the average action video game. We specifically looked for "engagement decline" in three ways. First, we observed little difference in learning across graduate students (who are presumably more acclimatized to boring repetitive tasks) than undergraduates. Second, as shown in Figure 5, we saw no evidence of deterioration in performance for the video game task over the course of the experiment, as would be expected if engagement were declining. Finally, Figure 5 shows rapid improvements in performance during the first five video gaming sessions, but this does not result in any transfer of learning across the first two object discrimination sessions, as shown in Figure 6. Certainly, our experiment should not be taken as conclusive evidence that gamification cannot play a role in visual rehabilitation for prostheses users, but our results do suggest that developing successful gamification protocols may prove challenging.

Another possible reason why we did not witness learning transfer might have been that our "Fruit Ninja" game did not contain enough elements of an action video game. Some categories of video games appear to enhance perceptual skills more than others. It appears that the largest contributing factors are the speed and accuracy with which decisions need to be made in order to succeed in the game (Achtman, Green, & Bavelier, 2008). First- and third-person shooter games contain high-stakes, fast-paced, and unpredictable events and the need to "quickly aim at targets," which requires precise interactions between visuomotor and cognitive systems to succeed. Video games in categories such as sports and racing (e.g., FIFA, NBA2k), strategy (e.g., Roller Coaster Tycoon, Sims), and puzzle or card games can be visually complex, require quick reallocation of attention, and demand the tracking of multiple objects but do not typically lead to increased visual skills. This is likely because the visual demands of these games are not as intense as those required by fast-paced action video games.

Our "Fruit Ninja" game was designed to capture as many of the elements thought to elicit fast transferable learning as possible while still being playable.

Unfortunately, pilot testing revealed that it was nearly impossible to play a filtered version of any commercially available action video game (Mortal Kombat II) in a way that was engaging or comprehensible to participants. While we tried to develop a game that included as many critical elements as possible, it is possible that a more complex game might have been more successful in generating transferable learning.

Effects of student status

The Gaming group consistently outperformed the Non-Gaming group even in pretests that were carried out prior to any experimental difference between the two groups. This is likely due to the students' graduate status. Participants in the Non-Gaming group were only required to come into the lab for 11 sessions, as opposed to the Gaming group, which came in for 36 sessions. As a consequence, the Gaming group ended up being more heavily composed of graduate students in the University of Washington psychology department (8 graduate students out of 10 in the Gaming group vs. 2 out of 10 in the Non-Gaming group). Graduate students tended to outperform undergraduate students, regardless of experimental group. Graduate students in psychology would have been well-aware of the importance of staying attentive during an experiment, despite a dull and repetitive task (Eyman, 1966), and likely were more motivated to perform well (Galanter, 1962).

It is important to note that although this selection bias resulted in large performance differences, it did not result in differences in learning, suggesting that our primary measure of interest—*learning*—is relatively robust to selection biases.

Conclusions

Our method of presenting conflicting spatial frequency and orientation information to each eye represents a novel way to study the distortions caused by electronic sight restoration technologies, by creating "virtual patients." Participants improved over sessions for both the discrimination task and the video game, but the failure of learning to transfer from the video game to the object discrimination task suggests that the learning was highly specific to each task. In contrast, although our transfer of learning studies (see Appendix) were individually underpowered, they all seemed to show similar amounts of learning as the training stimulus, suggesting that learning transferred across different types of filtering. This pattern of results is consistent with the idea that participants were simply improving in their ability to discriminate "garbled" objects in a highly task-specific manner. Overall, our results suggest that rehabilitation protocols for sight recovery technologies may be somewhat limited in utility and may prove most useful when focused on solving specific visual tasks.

Keywords: visual prostheses, sight restoration, perceptual learning, video game training, plasticity

Acknowledgments

Supported by NIH grant EY031312.

Commercial relationships: none.
Corresponding author: Ione Fine.
Email: ionefine@uw.edu.
Address: Department of Psychology, University of Washington, WA 98195-1525, USA.

References

AbbVie. (2022). RST-001 Phase I/II Trial for Advanced Retinitis Pigmentosa (NCT02556736), https://www.abbvieclinicaltrials.com/study/?id=RST-001-CP-0001.

Achtman, R. L., Green, C. S., & Bavelier, D. (2008). Video games as a tool to train visual skills. *Restorative Neurology and Neuroscience, 26*(4–5), 435–446.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*(6631), 401–406.

Antolik, J., Sabatier, Q., Galle, C., Fregnac, Y., & Benosman, R. (2021). Assessment of optogenetically-driven strategies for prosthetic restoration of cortical vision in large-scale neural simulation of V1. *Scientific Reports, 11*(1), 10783.

Ayton, L. N., Blamey, P. J., Guymer, R. H., Luu, C. D., Nayagam, D. A., Sinclair, N. C., . . . Bionic Vision Australia Research, C. (2014). First-in-human trial of a novel suprachoroidal retinal prosthesis. *PLoS One, 9*(12), e115239.

Ayton, L. N., Rizzo, J. F., III, Bailey, I. L., Colenbrander, A., Dagnelie, G., Geruschat, D. R., . . . Taskforce, H. I. (2020). Harmonization of outcomes and vision endpoints in vision restoration trials: Recommendations from the International HOVER Taskforce. *Translational Vision Science & Technology, 9*(8), 25.

Bach, M. (1996). FrACT-Landolt-Vision. *Optometry and Vision Science, 73*(1), 49–53.

Bach, M. (2007). The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis.

Graefe's Archive for Clinical and Experimental Ophthalmology, 245(7), 965–971.

Bavelier, D., & Green, C. S. (2019). Enhancing attentional control: Lessons from action video games. *Neuron, 104*(1), 147–163.

Beauchamp, M. S., Oswalt, D., Sun, P., Foster, B. L., Magnotti, J. F., Niketeghad, S., . . . Yoshor, D. (2020). Dynamic stimulation of visual cortex produces form vision in sighted and blind humans. *Cell, 181*(4), 774–783.e775.

Bosking, W. H., Sun, P., Ozker, M., Pei, X., Foster, B. L., Beauchamp, M. S., . . . Yoshor, D. (2017). Saturation in phosphene size with increasing current levels delivered to human visual cortex. *Journal of Neuroscience, 37*(30), 7188–7197.

Campbell, R., Evans, M., Tucker, M., Quilty, B., Dieppe, P., & Donovan, J. L. (2001). Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. *Journal of Epidemiology and Community Health, 55*(2), 132–138.

Caspi, A., Roy, A., Dorn, J. D., & Greenberg, R. J. (2017). Retinotopic to spatiotopic mapping in blind patients implanted with the Argus II retinal prosthesis. *Investigative Ophthalmology & Visual Science, 58*(1), 119–127.

Castel, A. D., Pratt, J., & Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica, 119*(2), 217–230.

Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *Journal of Neurophysiology, 78*(6), 2889–2894.

da Cruz, L., Dorn, J. D., Humayun, M. S., Dagnelie, G., Handa, J., Barale, P. O., . . . Argus, I. I. S. G. (2016). Five-year safety and performance results from the Argus II retinal prosthesis system clinical trial. *Ophthalmology, 123*(10), 2248–2254.

Dye, M. W., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science, 18*(6), 321–326.

Erickson-Davis, C., & Korzybska, H. (2021). What do blind people "see" with retinal prostheses? Observations and qualitative reports of epiretinal implant users. *PLOS ONE, 16*(2), e0229189.

Esquenazi, R. B., Meier, K., Beyeler, M., Boynton, G. M., & Fine, I. (2021). Learning to see again: Perceptual learning of simulated abnormal on- off-cell population responses in sighted individuals. *Journal of Vision, 21*(13), 10, doi:10.1167/jov.21.13.10.

Essery, R., Geraghty, A. W., Kirby, S., & Yardley, L. (2017). Predictors of adherence to home-based physical therapies: A systematic review. *Disability and Rehabilitation, 39*(6), 519–534.

Eyman, R. K. (1967). The effect of sophistication on ratio- and discriminative scales. *Am J Psychol, 80*(4), 520–540.

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*(2), 154–160.

Fernandez, E., Alfaro, A., Soto-Sanchez, C., Gonzalez-Lopez, P., Lozano, A. M., Pena, S., . . . Normann, R. A. (2021). Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex. *Journal of Clinical Investigation, 131*(23).

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning tasks: A review. *Journal of Vision, 2*(2), 190–203.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature, 287*(5777), 43–44.

Fu, E., Wang, T., Li, J., Yu, M., & Yan, X. (2022). Video game treatment of amblyopia. *Survey of Ophthalmology, 67*(3), 830–841.

Fujikado, T., Kamei, M., Sakaguchi, H., Kanda, H., Endo, T., Hirota, M., . . . Nishida, K. (2016). One-year outcome of 49-channel suprachoroidal-transretinal stimulation prosthesis in patients with advanced retinitis pigmentosa. *Investigative Ophthalmology & Visual Science, 57*(14), 6147–6157.

Galanter, E. (1962). *Contemporary psychophysics*. New York, NY: Holt.

Gambacorta, C., Nahum, M., Vedamurthy, I., Bayliss, J., Jordan, J., Bavelier, D., . . . Levi, D. M. (2018). An action video game for the treatment of amblyopia in children: A feasibility study. *Vision Research, 148*, 1–14.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534–537.

Green, C. S., & Bavelier, D. (2019). Corrigendum: Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 30*(12), 1790.

Green, C. S., Li, R., & Bavelier, D. (2010). Perceptual learning during action video game playing. *Topics in Cognitive Science, 2*(2), 202–216.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Haim, M. (2002). Epidemiology of retinitis pigmentosa in Denmark. *Acta Ophthalmologica,* 233, 1–34.

Hornig, R., Dapper, M., Le Joliff, E., Hill, R., Ishaque, K., Posch, C., . . . Picaud, S. (2017). Pixium vision: First clinical results and innovative developments. In V. P. Gabel (Ed.), *Artificial vision* (pp. 99–113). Cham, Switzerland: Springer International Publishing.

Hussain, Z., Astle, A. T., Webb, B. S., & McGraw, P. V. (2014). The challenges of developing a contrast-based video game for treatment of amblyopia. *Frontiers in Psychology, 5*, 1210.

Jeter, P. E., Dosher, B. A., Petrov, A., & Lu, Z. L. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision, 9*(3),1: 1–13.

John, M. C., Quinn, J., Hu, M. L., Cehajic-Kapetanovic, J., & Xue, K. (2022). Gene-agnostic therapeutic approaches for inherited retinal degenerations. *Frontiers in Molecular Neuroscience, 15*, 1068185.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 88*(11), 4966–4970.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*(6443), 250–252.

Kashani, A. H. (2022). Stem cell-derived retinal pigment epithelium transplantation in age-related macular degeneration: Recent advances and challenges. *Current Opinion in Ophthalmology, 33*(3), 211–218.

Korvasová, K., Grani, F., Voldřich, M., Peco, R. L., Berling, D., Calvo, M. V., . . . Antolík, Ján (2025). Contributed Talks I: Recruiting native visual representations in visual cortex for electrode array based vision restoration. *Journal of Vision, 25*(5), 10.

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience, 12*(5), 549–551.

Godot Engine Contributors, (2020). Godot Engine 3.2.2. https://godotengine.org.

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.

Matthews, N., Liu, Z., Geesaman, B. J., & Qian, N. (1999). Perceptual learning on orientation and direction discrimination. *Vision Research, 39*(22), 3692–3701.

McClements, M. E., Staurenghi, F., MacLaren, R. E., & Cehajic-Kapetanovic, J. (2020). Optogenetic gene therapy for the degenerate retina: Recent advances. *Frontiers in Neuroscience, 14*, 570909.

Mohanty, S. K., Mahapatra, S., Batabyal, S., Carlson, M., Kanungo, G., Ayyagari, A., . . . Mahajan, V. B. (2025). A synthetic opsin restores vision in patients with severe retinal degeneration. *Molecular Therapy, 33*(5), 2279–2290.

Ni, A. M., & Maunsell, J. H. (2010). Microstimulation reveals limits in detecting different signals from a local cortical region. *Current Biology, 20*(9), 824–828.

Oei, A. C., & Patterson, M. D. (2015). Enhancing perceptual and attentional skills requires common demands between the action video games and transfer tasks. *Frontiers in Psychology, 6*, 113.

Ohlschlager, S., & Vo, M. L. (2017). SCEGRAM: An image database for semantic and syntactic inconsistencies in scenes. *Behavior Research Methods, 49*(5), 1780–1791.

Palanker, D., Le Mer, Y., Mohand-Said, S., & Sahel, J. A. (2022). Simultaneous perception of prosthetic and natural vision in AMD patients. *Nature Communications, 13*(1), 513.

Provansal, M., Marazova, K., Sahel, J. A., & Picaud, S. (2022). Vision restoration by optogenetic therapy and developments toward sonogenetic therapy. *Translational Vision Science & Technology, 11*(1), 18.

R Core Team, (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rizzo, S., Belting, C., Cinelli, L., Allegrini, L., Genovesi-Ebert, F., Barca, F., . . . di Bartolo, E. (2014). The Argus II Retinal Prosthesis: 12-month outcomes from a single-study center. *American Journal of Ophthalmology, 157*(6), 1282–1290.

Sahel, J. A., Boulanger-Scemama, E., Pagot, C., Arleo, A., Galluppi, F., Martel, J. N., . . . Roska, B. (2021). Partial recovery of visual function in a blind patient after optogenetic therapy. *Nature Medicine, 27*(7), 1223–1229.

Saunders, A. L., Williams, C. E., Heriot, W., Briggs, R., Yeoh, J., Nayagam, D. A., . . . Allen, P. J. (2014). Development of a surgical procedure for implantation of a prototype suprachoroidal retinal prosthesis. *Clinical & Experimental Ophthalmology, 42*(7), 665–674.

Shen, Y. (2020). Stem cell therapies for retinal diseases: From bench to bedside. *Journal of Molecular Medicine, 98*(10), 1347–1368.

Simunovic, M. P., Shen, W., Lin, J. Y., Protti, D. A., Lisowski, L., & Gillies, M. C. (2019). Optogenetic approaches to vision restoration. *Experimental Eye Research, 178*, 15–26.

Smith, W., Assink, J., Klein, R., Mitchell, P., Klaver, C. C., Klein, B. E., . . . de Jong, P. T. (2001). Risk factors for age-related macular degeneration: Pooled findings from three continents. *Ophthalmology, 108*(4), 697–704.

Stingl, K., Bartz-Schmidt, K. U., Besch, D., Chee, C. K., Cottriall, C. L., Gekeler, F., . . . Zrenner, E. (2015). Subretinal Visual Implant Alpha IMS—Clinical trial interim report. *Vision Research, 111*(Pt. B), 149–160.

Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: The "Catch the Spies" task. *Cognitive Development, 20*(3), 373–387.

Troyk, P. R. (2017). The Intracortical Visual Prosthesis Project. In V. P. Gabel (Ed.), *Artificial vision: A Clinical Guide* (pp. 203–214). Cham: Springer International Publishing.

Van Gelder, R. N. (2015). Photochemical approaches to vision restoration. *Vision Research, 111*(Pt. B), 134–141.

Varela, M. D., de Guimaraes, T. A. C., Georgiou, M., & Michaelides, M. (2022). Leber congenital amaurosis/early-onset severe retinal dystrophy: current management and clinical trials. *British Journal of Ophthalmology, 106*(4), 445–451.

Wang, H., Yang, Y., Liu, J., & Qian, L. (2021). Direct cell reprogramming: approaches, mechanisms and progress. *Nature Reviews Molecular Cell Biology, 22*(6), 410–424.

Wood, E. H., Tang, P. H., De la Huerta, I., Korot, E., Muscat, S., Palanker, D. A., . . . Williams, G. A. (2019). Stem cell therapies, gene-based therapies, optogenetics and retinal prosthetics: Current state and implications for the future. *Retina, 39*(5), 820–835.

Zarbin, M., Sugino, I., & Townes-Anderson, E. (2019). Concise review: Update on retinal pigment epithelium transplantation for age-related macular degeneration. *STEM CELLS Translational Medicine, 8*(5), 466–477.