

# Resultado de tareas visuales complejas utilizando prótesis de visión simulada mediante gafas de realidad aumentada.

**Elton Ho**

Departamento de Física, Universidad de Stanford, Stanford, CA, USA

Laboratorio de Física experimental Hansen, Universidad de Stanford, Stanford, CA, USA

**Jack Boffa**

Laboratorio de Física experimental Hansen, Universidad de Stanford, Stanford, CA, USA

**Daniel Palanker**

Laboratorio de Física experimental Hansen, Universidad de Stanford, Stanford, CA, USA

Departamento de oftalmología, Universidad de Stanford, Stanford, CA, USA

Las prótesis subretinianas fotovoltaicas están diseñadas para restaurar la visión central en pacientes con degeneración asociada con la edad (DMAE). Hemos investigado la utilidad de las prótesis de visión central para tareas visuales complejas utilizando gafas de realidad aumentada (AR) que simulaban una agudeza, contraste y campo visual reducido. Las gafas AR con campo visual central bloqueado de 20° incluían una cámara de video integrado y un software que ajusta la calidad de imagen de acuerdo a tres parámetros establecidos por el usuario: Resolución, correspondiendo al tamaño de pixel equivalente a un implante; campo de visión, correspondiendo al tamaño del implante; y el número niveles de escalas de grises. El video procesado en tiempo real era transmitido en una imagen delante del ojo derecho. 19 participantes sanos fueron reclutados para completar tareas visuales incluyendo tablas optométricas, lecturas de enunciados y reconocimiento facial. En las tablas optométricas, la agudeza de letras excedió el límite de muestreo de píxeles en 0,2 logMar. La velocidad de lectura disminuyó tras incrementar el tamaño del pixel y con un campo de visión reducido (7°-12°). En la tarea de reconocimiento facial (elección forzada en cuatro direcciones, 5° de tamaño angular) los participantes identificaron las caras con una precisión >75%, incluso con píxeles de 100 µm y solo dos niveles de escala de grises. Con píxeles de 60 µm y ocho niveles de escala de grises la precisión superó el 97%. Sujetos con prótesis visuales simuladas resultaron ligeramente mejor que el límite de muestreo en el examen con tablas optométricas, y fueron más precisos en el reconocimiento de caras, incluso con una resolución de 100 µm/pixel. Estos resultados muestran la viabilidad de la lectura y el reconocimiento de caras mediante el uso de prótesis de visión central incluso con píxeles de 100 µm, y el rendimiento mejora aún más con píxeles más pequeños.

## Introducción

La degeneración macular asociada con la edad (DMAE) es la principal causa de deterioro visual intratable. Con la actual prevalencia mundial de 8,7%, se estima que la DMAE afecta casi 200 millones de personas en 2020, y su prevalencia está aumentando con la edad poblacional (Friedman, Tomany, McCarty, & De Jong, 2004; Wong et al., 2014). Pacientes con atrofia avanzada DMAE (actualmente con una prevalencia del 15 en países de occidente (Friedman et al., 2004; Wong et al., 2014) ) sufren de la pérdida de los fotorreceptores en la mácula, afectando la visión central. Aunque se pierde la visión de alta resolución, los pacientes todavía pueden usar su visión periférica preservada y normalmente conservan la agudeza en no peor que 20/400. Por tanto, la restauración de la visión central puede resultar útil siempre que la agudeza visual recuperada excede los niveles naturales.

En el ojo sano, los fotorreceptores transforman la luz incidente en pulsos químicos y eléctricos. Las señales neurales resultantes son procesadas por las células bipolares y otras neuronas que no presentan impulsos en las capas nucleares internas (INL) y avanzan hacia las células ganglionares de la retina (RGC), que generan potenciales de acción que se propagan a través del nervio óptico al cerebro. La pérdida de fotorreceptores en enfermedades degenerativas de la retina altera el proceso de fototransducción inicial, mientras que la red retiniana restante permanece intacta, aunque con algunos recableados (Humayun et al., 1999; Kim et al., 2002; Mazzoni, Novelli y Strettoi, 2008).

Se están desarrollando múltiples enfoques para abordar la pérdida de visión en la degeneración retiniana (Scholl et al., 2016), incluida la terapia génica (Sengillo,

Justus, Tsai, Cabral y Tsang, 2016), el trasplante de células (Lorach et al., 2019) ; Seiler et al., 2008), optogenética (Barrett, Berlinguer Palmi y Degenaar, 2014) e implantes electrónicos. En este último caso, se coloca una serie de electrodos en el punto de estimulación, como la retina (D. Palanker & Goetz, 2018), el nervio óptico (Veraart, Wanet-Defalque, Gerard, Vanlierde y Delbeke, 2003), el núcleo lateral geniculado (LGN; Nguyen et al., 2016), o la corteza visual primaria (Lewis, Ackland, Lowery y Rosenfeld, 2015). Se inyecta una corriente eléctrica en el tejido para estimular las células y de ese modo lograr la percepción visual. Tras la activación del electrodo, los pacientes informan que perciben "puntos brillantes", denominados fosfenos (Humayun et al., 2012; Stingl et al., 2015). La cantidad de electrodos limita la cantidad de información que se puede entregar y la densidad de los electrodos restringe la resolución más alta posible. En los estudios en animales con prótesis retinianas fotovoltáicas, demostramos que la agudeza de la rejilla coincide con el paso de píxeles de 55 (Ho, Lorach, Huang, et al., 2018) y 75  $\mu\text{m}$  (Lorach, Goetz, Smith, et al., 2015). Un ensayo clínico reciente de dichos implantes (PRIMA, de Pixium Vision) que tienen píxeles de 100  $\mu\text{m}$  también demostró que la agudeza visual protésica en pacientes con DMAE está solo entre un 10% y un 30% por debajo del límite de muestreo de 20/420 para el tamaño de píxel actual (D.V. Palanker et al., 2019).

El implante PRIMA estimula la primera capa de neuronas después de los fotorreceptores (INL) y, por lo tanto, las respuestas retinianas mediadas por redes provocadas conservan muchas características del procesamiento natural de la señal, incluida la fusión de parpadeo a altas frecuencias (>20 Hz; Lorach, Goetz, Mandel, et al., 2015; Lorach, Goetz, Smith, et al., 2015), la adaptación a imágenes estáticas (Stingl et al., 2013), la organización antagonista centro-envolvente de campos receptivos con suma lineal y no lineal de sus subunidades (Ho et al., 2017). Los pacientes con el implante pueden percibir líneas tan delgadas como el tamaño de los píxeles en la retina (es decir, un solo píxel de ancho) e identificar letras con el espacio mínimo en la letra C de 1,1 a 1,3 píxeles (D. V. Palanker et al., 2019).

Dado que los pacientes con DMAE conservan la visión periférica, tienen pocos problemas con la deambulación. Sin embargo, la lectura deficiente y el reconocimiento facial plantean desafíos importantes en la vida diaria (Mitchell y Bradley, 2006). Para evaluar la resolución espacial, el número de niveles de escala de grises y el tamaño del implante necesario para estas tareas visuales, simulamos la visión central protésica utilizando gafas de realidad aumentada con una cámara. La visión protésica se imitó mediante la reducción controlada de la resolución espacial, el contraste y el campo visual de las imágenes proyectadas en la pantalla incorporada. Aquí, investigamos qué tan bien los sujetos sanos pueden realizar tareas visuales complejas, incluida la lectura y el reconocimiento facial, bajo varios niveles de degradación de la imagen. Los resultados de este estudio predicen los mejores resultados clínicos posibles, ya que la visión protésica en pacientes con degeneración de la retina es

probablemente peor que la visión natural pixelada con contraste reducido.

En el pasado se realizaron estudios psicofísicos de visión protésica simulada, pero consideramos que esos resultados son insuficientes para predecir los resultados con nuestro implante actual. Con el implante subretiniano fotovoltáico para la restauración de la visión central en pacientes con DMAE, la simulación requiere las siguientes especificaciones: (a) densidad de píxeles >100 píxeles/ $\text{mm}^2$ , (b) sin espacios entre fosfenos, (c) campo visual en el rango de  $7^\circ$ - $10^\circ$  y (d) se permite la exploración ocular. Dado que los estudios anteriores no abordaron estas especificaciones, realizamos un estudio psicofísico para evaluar los límites del rendimiento visual del sistema PRIMA y establecer las expectativas para los próximos ensayos clínicos.

## Métodos

### Sújetos

19 sujetos (edades 18-74 años), todos reclutados de la plantilla de la Universidad de Stanford, quienes concedieron su permiso firmado y participaron en el actual estudio. Todos los individuos se habían autodiagnosticado con visión normal, además, su agudeza visual fue verificada tanto por el test de Landolt C y optotipos ETDRS antes de los experimentos. Para tareas de lectura complejas, los sujetos debían poseer una competencia cercana a la de un nativo o nativa en inglés. Todos los sujetos presentaban una experiencia previa limitada o nula con gafas de realidad virtual-aumentada (AR). El estudio fue aprobado por el grupo IRB de Stanford para la investigación con sujetos humanos y dirigido de acuerdo a las directrices institucionales, siguiendo los postulados de la Declaración de Helsinki.

### Montaje experimental

El equipo experimental incluía dos partes: un sistema de presentación de estímulos y gafas AR con la pantalla frontal y una unidad de procesamiento de imágenes (Figura 1).

El sistema de producción de estímulos estaba compuesto por un monitor de 24" (ASUS VS248H-P) controlado por un ordenador portátil (Thinkpad 25, Lenovo) utilizando un software custom llamado PsychToolbox (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli & Vision, 1997) en MatLab. Este sistema fue utilizado para mostrar los estímulos y registrar las respuestas de los sujetos (como la precisión y el tiempo empleado) a través de la contribución del experimentador. El monitor fue situado a 30" de una mentonera, donde los sujetos colocarían la cabeza durante un experimento. El monitor tenía una resolución de 2400 x 1350 píxeles, correspondiendo a 90.4 píxeles por grado (ppd) de ángulo

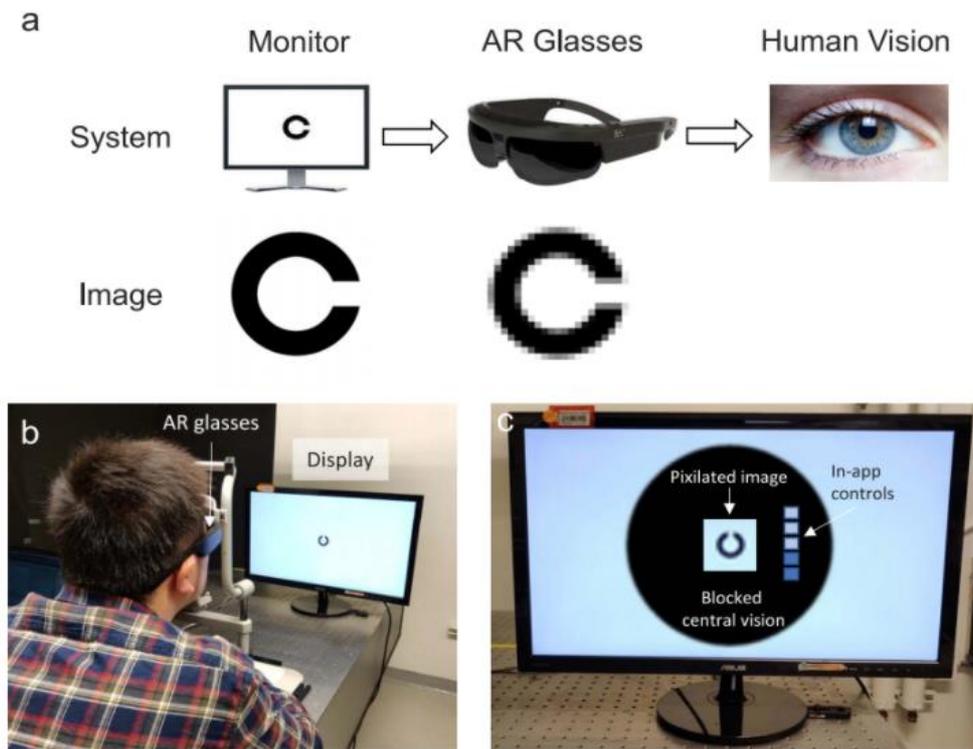


Figura 1. Equipo experimental. (a) Esquema del montaje experimental. Las imágenes de alta resolución se presentan un monitor. La parte frontal de la cámara de las gafas de realidad aumentada (AR) capturan la emisión de video. El software personalizado precargado en las gafas AR ajusta la calidad del video para imitar la visión protésica y la muestra en las gafas AR. (b) Un sujeto frente al aparato. (c) Ilustración de la visión a través de las gafas AR.

visual. La cámara (4 MP) montada en la parte frontal de las gafas AR (ODG R 7, Osterhout Design Group, San Francisco, CA) captura una transmisión de video en vivo. La ampliación de la cámara se configuró para que coincidiera con el tamaño angular de la visión natural. Luego, los datos se procesan con una aplicación personalizada en Android a tiempo real de acuerdo con tres parámetros definidos por el usuario: pixelación (equivalente a un tamaño de píxel de 30 a 100  $\mu\text{m}$  en la retina), número de niveles de escala de grises (2 a 256) y campo de visión (FOV; 7°-12°). El video resultante se presentó en la pantalla de las gafas (especificaciones: 30° FOV, 720 p, 80 fps). La latencia entre la cámara y la pantalla fue mínima debido al rápido procesamiento de video. En un sistema de pantalla AR típico, la pantalla integrada es transparente, por lo que la información visual presentada puede fusionarse con el fondo de transferencia (de ahí, realidad "aumentada").

Para imitar la pérdida de visión en pacientes con DMAE, se bloqueó un área de los anteojos correspondiente a 20° de la visión central con cinta negra opaca para ambos ojos. En esta región, solo la pantalla integrada era visible, mientras que fuera de esa región, solo estaba presente la visión periférica natural (Figura 1c). Aquí solo evaluamos la visión protésica monocular, por lo que la pantalla solo se encendió para el ojo derecho, que por cierto correspondía al ojo dominante de todos los sujetos. La información visual para objetivos pequeños (Landolt C, identificación de letras y reconocimiento facial) abarcaba un máximo de 5°, por lo que los sujetos usaban exclusivamente la pantalla electrónica y no podían beneficiarse moviendo sus ojos fuera del área oscurecida. De manera similar, en las pruebas de lectura de oraciones

con fuentes de 1° o menos, todo el texto se mostró detrás de la máscara. En cuanto a las oraciones con tamaños de letra más grandes, aunque en teoría los sujetos podrían haberse asomado al exterior, la mayor parte de la oración seguía oscurecida. Se instruyó a los sujetos para que vocalizaran las oraciones de manera secuencial, por lo que al mirar a las pocas letras del final de la oración no ayudaría con la tarea. Por lo tanto, los participantes prácticamente solo leyeron a través de la pantalla electrónica, lo que se confirmó mediante entrevistas posteriores.

El procesamiento de video fue realizado con la biblioteca OpenCV, y el flujo de trabajo fue el siguiente: Un fotograma del video fue recortado para coincidir con el FOV deseado. Luego, el fotograma se convirtió a escala de grises, se redujo y se volvió a aumentar al tamaño de la imagen original, lo que resultó en una imagen en escala de grises pixelada y compacta. Usamos la interpolación por el vecino más cercano para la transformación de imágenes en Android. La pixelación aquí coincidió con el tamaño de píxel deseado en la retina, por ejemplo, píxeles de 100  $\mu\text{m}$  subtienden 0.35° en la retina humana). El valor de color gris para cada píxel se redondeó luego al 255/(n-1) más cercano, donde  $n$  es el número de niveles de la escala de grises.

Los sujetos fueron instruidos para llevar gafas AR y ajustar el tamaño del píxel, los niveles de escala de grises, y el FOV usando los controles dentro de la misma aplicación. Para familiarizar a nuestros individuos con la visión protésica simulada, se les dio instrucciones de mirar alrededor del laboratorio libremente por unos pocos minutos, y se les mostraron algunas imágenes de animales comunes, plantas y alimentos.

Tarea	Equivalencia tamaño pixel en la retina ( $\mu\text{m}$ )	Número de niveles de escalas de grises	FOV ( $^{\circ}$ )
Agudeza visual	30, 60, 100	2, 8	7
Lectura de enunciados	30, 60, 100	8	7, 12
Reconocimiento facial	60, 100	2, 4, 8	7

Tabla 1. Parámetros del procesamiento de imágenes usado para cada experimento.

## Procedimientos

Dirigimos experimentos diferentes: (a) agudeza visual en reconocimiento de letras, (b) lectura de enunciados, y (c) reconocimiento facial. Los parámetros para la visión protésica simulada se resumen en la Tabla 1. Se indicó a los sujetos que fijaran su visión central en el centro de la pantalla AR, pero se les permitió mover los ojos y la cabeza, si así lo deseaban. En todos los experimentos, los sujetos vocalizaron sus respuestas, que fueron grabadas y cronometradas por el experimentador. Normalmente, un conjunto completo de experimentos puede durar hasta 90 minutos. Si un sujeto se cansaba, se programaba una nueva sesión para las tareas restantes.

### Agudeza visual en reconocimiento de letras

Se pidió a los sujetos ( $n=13$  para 30 y 60  $\mu\text{m}$ ;  $n=19$  para visión natural y 100  $\mu\text{m}$  de tamaño de píxeles) que identificaran la orientación del Landolt C, mostrando uno a la vez. Si el sujeto pudo identificar al menos cuatro de cinco orientaciones del mismo tamaño, reducíamos el tamaño de la letra en 0.1 LogMAR unidades y se repetía. El mismo experimento fue dirigido con los optotipos ETDRS con fuente Sloan. El rasgo más pequeño de estos caracteres era un quinto del tamaño de la letra. El primero se evaluó la agudeza visual de los sujetos con visión normal o corregida a normal sin anteojos AR, y luego con visión protésica simulada. Como punto de comparación, también calculamos el límite de muestreo para cada tamaño de píxel protésico calculando su agudeza visual equivalente geométrica.

### Lectura de enunciados

Se pidió a los sujetos ( $n = 9$  para enunciados simples;  $n = 10$  para enunciados complejos) que leyeren en voz alta los enunciados que se les mostraba tan rápido como pudieran siguiendo el protocolo standard MNREAD (<http://legge.psych.umn.edu/mnread-set>). Se presentaron 3 líneas de texto con fuente Arial con 20 caracteres por línea aproximadamente. Una nueva frase con un tamaño de fuente menor (-0.1 LogMAR) se les mostró después de pronunciar correctamente ( $\leq 2$  errores por oración). El tamaño de fuente fue medido como el ángulo visual entre la parte visual entre la parte superior de la letra “k” y la parte inferior de la letra “p”. Entre las oraciones se mostró una cruz de fijación en el centro de la pantalla durante 2 s para volver a centrar la visión de los sujetos. Los sujetos

fueron evaluados primero con su visión binocular normal/corregida, y luego con visión protésica simulada con diferentes tamaños de píxel y FOV. Se utilizaron 8 niveles de escalas de grises para cumplir con las expectativas máximas de los estudios con roedores anteriormente presentados (Ho, Lorach, Goetz, et al., 2018) y los resultados con el implante Alpha IMS (Stingl et al., 2015). La velocidad de lectura (en palabras por minuto o PPM) para cada oración se registró en el software. Evaluamos el rendimiento de lectura con tres parámetros claves: precisión de lectura (RA, oración resoluble más pequeña, velocidad máxima de lectura (MRS) y tamaño de impresión crítico (CPS, tamaño de fuente más pequeño en el que se alcanza el 90% de MRS).

Los textos utilizados se pueden clasificar en oraciones simples y complejas. Las oraciones simples se compusieron internamente de acuerdo con el protocolo MNREAD o se tomaron de la aplicación MNREAD para iPad 2017 (<https://itunes.apple.com/us/app/mn-read/id1196638274?ls=1&mt=8>). Las oraciones complejas se seleccionaron Manually Annotated Sub-Corpus (MASC) del Open American National Corpus (OANC) (<http://www.anc.org/>) con tres criterios: número de caracteres entre 55 y 70, longitud promedio de palabra entre 5.5 y 6.5, y oraciones que se pueda segmentar en tres líneas de longitud similar. Por lo general, las oraciones MNREAD simples tienen un contexto independiente e involucran vocabulario a nivel de escuela primaria en EE.UU. (Por ejemplo, "Miró a su madre y le dijo que estaba muy feliz"), mientras que las oraciones complejas pueden incorporar más contexto con niveles avanzados de vocabulario (por ejemplo, "Un buen orden y limpieza contribuye a la seguridad y a obtener resultados fiables"). Solo se seleccionaron sujetos con un nivel de inglés nativo o casi nativo para la tarea de lectura compleja. Los resultados de oraciones simples y complejas se compararon mediante una prueba *t* de dos muestras.

### Reconocimiento Facial

A los sujetos ( $n = 19$  para 100  $\mu\text{m}$ ;  $n = 17$  para 60  $\mu\text{m}$ ) se les mostró un rostro adulto de referencia y debían elegir una cara entre cuatro que coincidiera con la identidad de la referencia lo más rápido posible (Figura 2a). Se registró la exactitud y el tiempo que les tomó para cada selección. Se llevó a cabo un conjunto de 10 intentos para cada combinación de parámetros, los cuales fueron presentados en un orden pseudoaleatorio para minimizar el efecto de memoria.

Se seleccionaron aleatoriamente imágenes no ocluidas de cabezas adultas de la base de datos Face Place (<http://www.tarrlab.org/>). La base de datos tiene una licencia CC BY-NC-SA 3.0 Unreported Licence. Para la misma identidad, un conjunto de imágenes incluía diferentes ángulos de visión y expresiones faciales, con el fondo recortado. Generalmente, las características más prominentes por encima del cuello eran visibles, incluido el peinado, el tono de piel y ambos ojos. Las imágenes se redimensionaron y recortaron a 5° x 5°. Se colocaron cinco imágenes en mosaico como se muestra en la Figura 2a, ocupando un campo visual de 16° x 16°. La imagen de referencia se colocó en el centro.

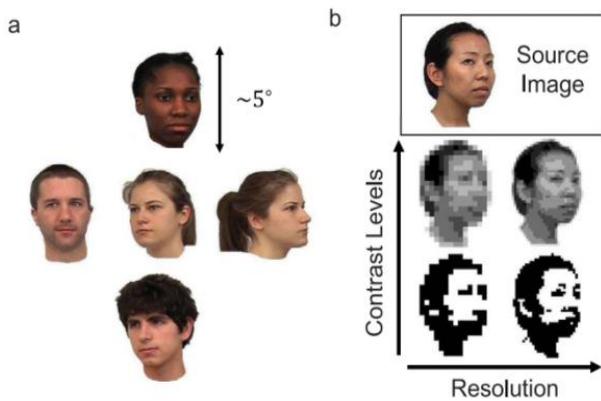


Figura 2. Tarea de reconocimiento facial. (a) Ejemplo de un conjunto de 5 rostros. Se pidió a los sujetos que escogieran el rostro que coincidiera con la identidad de la persona central. Cada rostro abarcaba aproximadamente 5° X 5°. (b) Efectos del número de niveles de escala de grises y la resolución de una imagen.

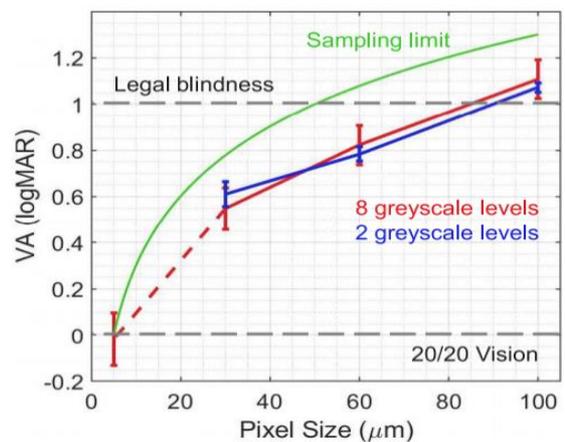


Figura 3. Resultados de agudeza visual ( $n = 13$  para píxeles de 30 y 60  $\mu\text{m}$ ;  $n = 19$  para visión natural y píxeles de 100  $\mu\text{m}$ ). El punto de medición más a la izquierda en 5  $\mu\text{m}$  indica agudeza visual (VA) para la visión estándar de los sujetos. Las barras de error están presentadas de acuerdo a SD.

## Resultados

### Agudeza Visual

Tanto con los optotipos ETDRS y el Landolt C, la VA mejoró con el tamaño de pixel reducido, tal y como se muestra en la figura 3. El punto de medición rojo más a la izquierda indica la agudeza visual normal o corregida de los sujetos, lo que sea mejor. La VA medida por ambos paradigmas de prueba concuerda entre sí. La prueba Landolt C arrojó una VA ligeramente mejor que los optotipos ETDRS en 0.05 logMAR, aunque apenas significativa (Suplemento, Figura 1). La disminución de los niveles de la escala de grises de 8 a 2 no afectó significativamente a la VA. Todas las VA medidas fueron al menos 0,2 logMAR mejores que el límite de muestreo calculado para cada tamaño de píxel. Esto podría atribuirse al sobremuestreo al escanear y a los sujetos que buscan diferencias entre letras submuestreadas. El reconocimiento de letras también fue mejor que el límite de muestreo y requirió solo alrededor de 3 píxeles por ancho de carácter para todos los tamaños de píxeles, de acuerdo con los 3 a 7 fosfenos por ancho de letra informados por otros estudios (Dagnelie, Barnett, Humayun y Thompson, 2006; Sommerhalder et al., 2003; Sommerhalder et al., 2004).

La mayor parte de los sujetos explicaron que cerca del límite no podían ver explícitamente la apertura de un Landolt C. Emplearon una estrategia en la que escanearon el objeto e identificaron el lado de la mancha que parpadeaba más, a través de la cual determinaban correctamente la orientación.

### Lectura de enunciados

Con un tamaño de píxel y un campo de visión limitados, la velocidad de lectura con visión protésica simulada (Figura 4, líneas verde y roja) fue mucho más

lenta que con visión natural sin obstrucciones (línea azul). La agudeza de lectura (RA) para la visión tanto natural como protésica coincidió con la agudeza de las letras correspondientes. A medida que el tamaño de la fuente aumentó por encima del umbral VA, la velocidad de lectura aumentó rápidamente hasta que se alcanzó la velocidad máxima de lectura (MRS) en el tamaño de impresión crítico (CPS). Un mayor aumento del tamaño de la fuente fue perjudicial, ya que menos palabras y letras podían caber en el campo de visión. Por ejemplo, una palabra de nueve letras con un tamaño de fuente de 1,58 (correspondiente a 1,58 de altura vertical y 0,788 de ancho horizontal asignados a cada letra) apenas puede caber en 78 de FOV. Para todos los tamaños de píxeles, el CPS fue aproximadamente el doble del RA, y el tamaño de fuente legible más pequeño fue de aproximadamente 2.5 píxeles por ancho de letra, un poco menos que la prueba de agudeza de letras y los informes anteriores. La discrepancia puede atribuirse al hecho de que, en las tareas de lectura, la pérdida de información letra por letra se compensa con pistas contextuales. En general, los píxeles más pequeños permitieron un muestreo más denso, lo que resultó en mejores RA, MRS y CPS. Mientras tanto, un mayor campo de visión no afectó significativamente a la RA, mientras que aumentó la velocidad de lectura con todos los tamaños de fuente superiores a CPS ( $t = 3.2$ ,  $p = 0,005$  para MRS). Los resultados numéricos se resumen en la Tabla 2.

Las tendencias generales con oraciones complejas fueron las mismas, aunque a menor velocidad (Figura 4b y Tabla 3). Sin embargo, el efecto del FOV sobre la MRS se volvió insignificante (ej.  $t = 1,45$ ,  $p = 0,156$  para 30  $\mu\text{m}$  / 12°). Contrariamente a la intuición, la RA y CPS fueron ligeramente mejores (más pequeños) para oraciones complejas que para simples, posiblemente debido a la predictibilidad de palabras en oraciones ricas en contexto.

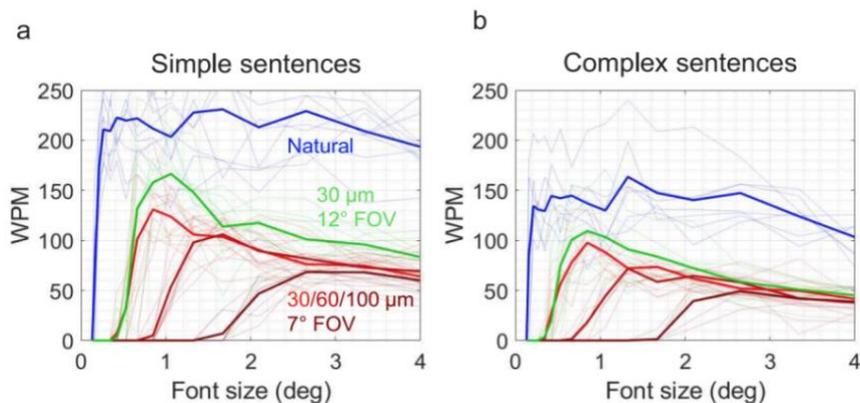


Figura 4. Velocidad de lectura de oraciones en palabras por minuto (PPM). (a) Oraciones sencillas. (b) Oraciones complejas. Las líneas atenuadas representan medidas individuales y las líneas en negrita representan la media de la población.

## Reconocimiento facial

Para todos los tamaños de píxeles y niveles de escala de grises, los sujetos podrían lograr una precisión superior al 75% en promedio, significativamente más alta que la elección aleatoria (25%; Figura 5). Si bien los rostros se reconocían casi instantáneamente con la visión natural, se necesitaron más de 5 segundos con la visión simulada, ya que se requería del escaneo para observar todos los rostros debido al campo visual limitado. El aumento del número de niveles de escala de grises y la reducción del tamaño de los píxeles mejoraron la precisión y el tiempo necesario para el reconocimiento facial. En la Figura 5c, los tiempos de respuesta se normalizan a los de píxeles de 100  $\mu\text{m}$  y ocho niveles de escala de grises. Una disminución en el tamaño de píxel de 100 a 60  $\mu\text{m}$  acortó el tiempo de respuesta en aproximadamente un 20% ( $p < 0.025$  para todos los niveles de escala de grises, prueba  $t$  de dos muestras). No hubo una diferencia significativa en la precisión entre píxeles 60 y 100  $\mu\text{m}$ .

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

Tabla 2. Agudeza de lectura (RA), velocidad de lectura máxima (MRS), y tamaño de impresión crítico (CPS) para lectura MNREAD

cantidades de píxeles (Dagnelie et al., 2006; Hayes et al., 2003; Irons et al., 2017; Shannon, 1992). Resultados clínicos recientes con prótesis subretinianas fotovoltaicas de píxeles de 100  $\mu\text{m}$  (PRIMA, Pixium Vision) confirmaron que la agudeza protésica en pacientes con DMAE, medida con la prueba Landolt C, casi coincide con el paso de píxeles (D. V. Palanker et al., 2019). Además, mediciones recientes con píxeles de 55  $\mu\text{m}$  en ratas demostraron que la agudeza de la rejilla también coincide con el tamaño de píxel de este tamaño (Ho, Lorach, Huang, et al., 2018). El desarrollo de electrodos tridimensionales permite píxeles aún más pequeños, que podrían proporcionar una resolución más alta en el futuro (Flores et al., 2018). Para evaluar los requisitos mínimos de un sistema de restauración de la visión central en pacientes con DMAE suficiente para lectura y reconocimiento facial, decidimos evaluar su rendimiento simulado en función de tres parámetros: tamaño de píxel, campo de visión (FOV) y número de niveles de escala de grises.

Los estudios anteriores con visión simulada utilizaron imágenes "fosfenizadas" (Chen, Hallum, Lovell y Suaning, 2005; Dagnelie et al., 2006; Thompson, Barnett, Humayun y Dagnelie, 2003). Se mostró un punto con un perfil plano o 2D-Gaussiano para simular un píxel activado, mientras que los puntos adyacentes se espaciaron de acuerdo con el tamaño de píxel del implante, lo que resultó en espacios oscuros entre los fosfenos simulados (Chen, Suaning, Morley y Lovell, 2009).

## Discusión

La agudeza en el reconocimiento de las letras y la velocidad de lectura son las métricas más comunes para evaluar la calidad de la visión, especialmente en pacientes con baja visión (Rubin, 2013). Agregamos una tarea de reconocimiento facial ya que es de alta prioridad para los pacientes con DMAE atrófica (Taylor, Hobby, Binns y Crabb, 2016). Muchos estudios de psicofísica con visión protésica simulada se diseñaron para investigar las capacidades potenciales de implantes con varias

Parámetros (píxeles/FOV)	Ra (°)	MRS (PPM)	CPS (°)
100 $\mu\text{m}/7^\circ$	$1.63 \pm 0.11$	$52 \pm 7$	$2.7 \pm 0.4$
60 $\mu\text{m}/7^\circ$	$0.71 \pm 0.11^*$	$77 \pm 10$	$1.7 \pm 0.5$
30 $\mu\text{m}/7^\circ$	$0.36 \pm 0.05^*$	$102 \pm 18$	$0.80 \pm 0.15$
30 $\mu\text{m}/12^\circ$	$0.37 \pm 0.09^*$	$121 \pm 33$	$0.87 \pm 0.25^*$
Natural	$0.14 \pm 0.01$	$173 \pm 34$	$0.32 \pm 0.09$

Tabla 3. Agudeza de lectura (RA), velocidad máxima de lectura (MRS), y punto crítico de tamaño de impresión (CPS) para la lectura de oraciones complejas utilizando visión pixelizada. *Notas:* Todos los errores son presentados como desviación estándar. El asterisco (\*) indica que  $p < 0.05$  (test  $t$  de dos muestras) comparado a oraciones sencillas con los mismos parámetros.

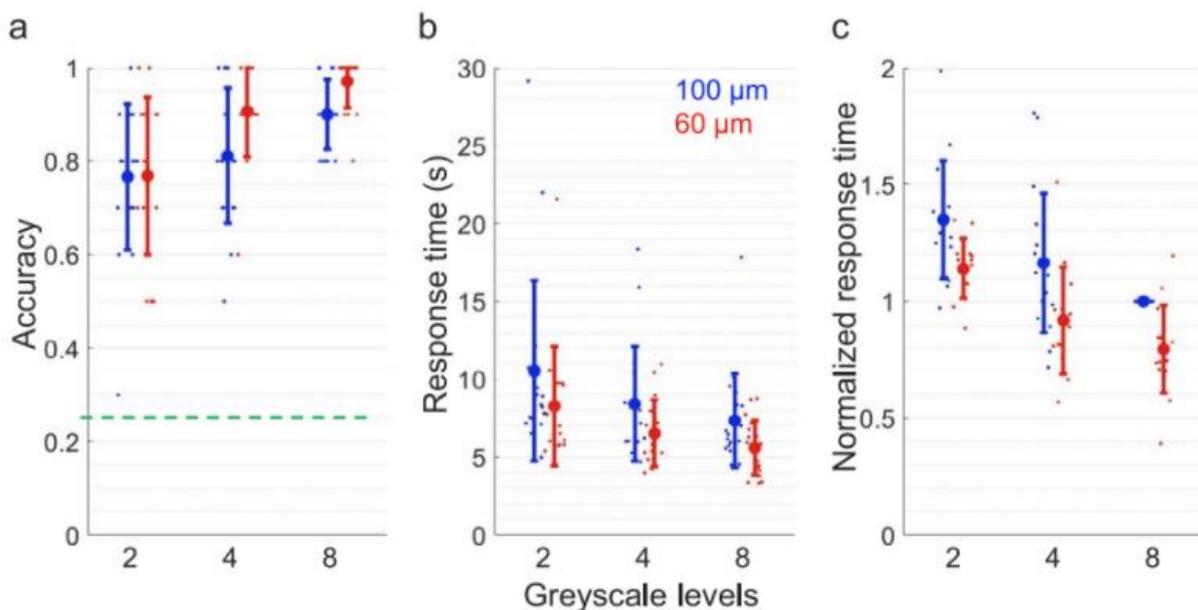


Figura 5. Reconocimiento facial. (a) Precisión. (b) Tiempo de respuesta. (c) Tiempo de respuesta normalizado a píxeles de 100  $\mu\text{m}$  y ocho niveles de escala de grises. Cada punto representa una medida independiente. Las barras de error se presentan en términos de SD.

Sin embargo, en el estudio clínico PRIMA (D. V. Palanker et al., 2019), al ver varios patrones de líneas, los pacientes anunciaron que percibían líneas continuas, en lugar de una fila de fosfenos desconectados. Por lo tanto, en nuestro estudio, usamos píxeles compactos, similares a los de un monitor de consumidor típico, sin espacios oscuros entre ellos.

Otra diferencia entre el estudio actual y los anteriores es la elección del FOV. Dado que otros implantes fueron diseñados para degeneraciones retinianas hereditarias que causan ceguera completa, su campo de visión funcional podría ser tan grande como 22° (Luo & da Cruz, 2016). Sin embargo, la atrofia geográfica rara vez supera los 4 mm de diámetro y, para evitar cualquier daño a la retina sana adyacente, el implante puede cubrir solo una parte del escotoma. Por lo tanto, es poco probable que los implantes subretinianos para DMAE excedan los 3 mm de ancho, lo que corresponde a aproximadamente 10° del ángulo visual. En el primer estudio de viabilidad, el tamaño del implante PRIMA es de 2 mm, correspondiente a aproximadamente 7° del campo visual (D. V. Palanker et al., 2019). Por lo tanto, estudiamos el efecto del campo de visión en la velocidad de lectura en el rango de 7° a 12°, mientras que toda la información visual para una agudeza en el reconocimiento de letras o tareas de reconocimiento facial estaba empaquetada dentro de los 5° del ángulo visual.

Cuando nuestros sujetos inicialmente no pudieron identificar la orientación del pequeño Landolt C, se les pidió que adivinaran sin que el experimentador afirmara la respuesta. Por lo general, los sujetos podrían detectar correctamente una o dos líneas adicionales de la tabla de agudeza, lo que explica que su desempeño exceda el límite de muestreo en aproximadamente 0.2 LogMAR, como se puede ver en la Figura 3. Esta estrategia se basa en escanear el objeto e identificar lo que es más oscuro o un tamaño parpadeante de la mancha abierta, que es suficiente para determinar la orientación del

Landolt C. Dicha estrategia se puede utilizar para otras tareas dentro de un pequeño grupo de patrones objetivo, como el reconocimiento de letras, pero es poco probable que ayude en la identificación de patrones y objetos desconocidos.

En el pasado se demostró repetidamente que la precisión del reconocimiento facial depende en gran medida de la resolución de la imagen, como se resume en (Irons et al., 2017). Con fosfenos de 16 x 16 por cara sobre un campo visual de 9,4° y 10 niveles de gris sin escanear, los sujetos pudieron diferenciar rostros con hasta un 84% de precisión (Chang, Kim, Shin y Park, 2012), uno de los más altos presentados. En otro estudio con fosfeno de 24 x 24 dentro de un FOV de 18°, la precisión fue del 65% y alcanzó el 88% con conjuntos de 32 x 32 (Wang et al., 2014). En el estudio actual, centrado en el modelado de pequeños implantes en la mácula central, utilizamos imágenes sustancialmente más pequeñas (que abarcan el rostro de 5° x 5°) con mayor densidad de píxeles, mientras que el número de píxeles por imagen era comparable a los de estudios anteriores. Descubrimos que se puede lograr una precisión casi perfecta en ocho niveles de escala de grises con píxeles de 60  $\mu\text{m}$ , correspondientes a una cuadrícula de 24 x 24.

Además de utilizar píxeles compactos y permitir el escaneo de la cabeza, otra explicación probable de un rendimiento mejorado es que cuando los rasgos faciales más prominentes se encuentran dentro de la fóvea (< 2 mm de diámetro), los sujetos pueden dedicar menos esfuerzo al escaneo y prestar más atención en la evaluación de los detalles faciales.

Curiosamente, la diferenciación facial por elección forzada en nuestro estudio requirió significativamente menos píxeles que el reconocimiento de objetos en un estudio anterior (Jung, Aloni, Yitzhaky y Peli, 2015). Con píxeles de 100  $\mu\text{m}$ , correspondientes a aproximadamente 200 píxeles por rostro, nuestros sujetos pudieron diferenciar rostros con una precisión de < 75%. Esto es mucho menos que los 560 píxeles

necesarios para reconocer objetos que cubren aproximadamente 10° del campo visual en un fondo ordenado. La diferencia podría deberse a una gran simplificación de la tarea cuando una referencia está disponible de inmediato, en comparación con nombrar un objeto entre un gran conjunto de opciones. Otra posibilidad es que los rostros puedan ser una clase de imágenes sorprendentemente fáciles de discernir. En un estudio que involucró diferentes clases de objetos y animales (Li, Hu, Chai y Peng, 2012), los sujetos demostraron una tasa de reconocimiento del .80% en todas las imágenes usando píxeles de 24 x 24. Sin embargo, con píxeles de 16 x 16, nadie podría reconocer un automóvil, pero el 90% pudo identificar un perro, lo que coincide con la precisión y los parámetros en nuestra tarea de reconocimiento facial. También es importante tener en cuenta que en nuestro estudio los rostros que presentamos estaban sobre un fondo blanco, mientras que con un fondo natural más desordenado, es posible que se necesiten de dos a tres veces más píxeles para lograr la misma precisión (Jung et al., 2015).

En conclusión, con la visión protésica simulada en gafas AR, los sujetos demostraron una agudeza en el reconocimiento de las letras que excedía ligeramente el límite de muestreo y una alta eficacia en el reconocimiento facial incluso con píxeles de 100  $\mu$ m píxeles. Estos resultados indican que los implantes fotovoltáicos subretinianos con píxeles de 100  $\mu$ m, actualmente disponibles para pruebas clínicas, pueden ser útiles para la lectura y el reconocimiento facial en pacientes que perdieron la visión central debido a la degeneración de la retina. Como era de esperar, los píxeles más pequeños mejoran significativamente el rendimiento visual y, por lo tanto, una mayor reducción del tamaño de los píxeles puede mejorar en gran medida los resultados en el futuro.

*Palabras Claves: Visión protésica, realidad aumentada, agudeza visual, lectura, reconocimiento facial.*

## Reconocimientos

Este trabajo fue subvencionado por los Institutos Nacionales de Salud (Grants R01-EY-018608, R01-EY-027786), el Departamento de Defensa (Grant W81XWH-15-1-0009), El Instituto de Neurociencia de Stanford, y Research to Prevent Blindness.

Imágenes de estímulo cortesía de Michael J. Tarr, Center for the Neural Basis of Cognition y Departamento de Psicología, Universidad Carnegie Mellon, <http://www.tarrlab.org/>. Financiación aportada por la NSF concesión 0339122.

Relaciones comerciales: DP ofrece información a Pixium Vision. Las patentes de DP correspondiente a las prótesis retinianas son propiedad de la Universidad de Stanford y

autorizadas para Pixium Vision. EH y JB certifican la nula competencia por intereses económicos.

Autor correspondiente: Elton Ho.

Email: eltonho@stanford.edu.

Dirección: Departamento de Física, Universidad de Stanford, Stanford, CA, USA; Laboratorio Experimental de física de Hansen, Universidad de Stanford, Stanford, CA, USA.

## Reconocimientos

1. Barrett, J. M., Berlinguer-Palmini, R., & Degenaar, P. (2014). Optogenetic approaches to retinal prosthesis. *Visual Neuroscience*, 31(4–5), 345–354.
2. Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
3. 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.
4. 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.
5. Chen, S. C., Suaning, G. J., Morley, J. W., & Lovell, N. H. (2009). Chen Simulating prosthetic vision: I. Visual models of phosphenes. *Vision Research*, 49(12), 1493–1506.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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  $\mu\text{m}$ . *Investigative Ophthalmology & Visual Science*, 59(9), 3977–3977.
12. 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.
  13. 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.
  14. 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.
  15. 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.
  16. Jung, J.-H., Aloni, D., Yitzhaky, Y., & Peli, E. (2015). Active confocal imaging for visual prostheses. *Vision Research*, 111, 182–196.
  17. Kim, S. Y., Saddy, 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.
  18. Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36(14), 1.
  19. 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.
  20. 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.
  21. 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>.
  22. 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>.
  23. 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.
  24. Luo, Y. H.-L., & da Cruz, L. (2016). The Argus II retinal prosthesis system. *Progress in Retinal and Eye Research*, 50, 89–107.
  25. 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.
  26. 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.
  27. 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.
  28. Palanker, D., & Goetz, G. (2018). Restoring sight with retinal prostheses. *Physics Today*, 71(7), 26–32.
  29. 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.
  30. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: *Transforming numbers into movies*. *Spatial Vision*, 10, 437–442.
  31. Rubin, G. S. (2013). Measuring reading performance. *Vision Research*, 90, 43–51.
  32. Scholl, H. P., Strauss, R. W., Singh, M. S., Dalkara, D., Roska, B., Picaud, S., & Sahel, J.-A. (2016). Emerging therapies for inherited retinal degeneration. *Science Translational Medicine*, 8(368), 368rv6.
  33. Seiler, M. J., Thomas, B. B., Chen, Z., Wu, R., Saddy, 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.
  34. 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).
  35. Shannon, R. V. (1992). A model of safe levels for electrical-stimulation. *IEEE Transactions on Biomedical Engineering*, 39(4), 424–426.
  36. 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.

37. 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.
38. 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>.
39. 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.
40. 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.
41. 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.
42. 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.
43. 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.
44. 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).