Holographic display system for restoration of sight to the blind

G A Goetz, Y Mandel, R Manivanh, D V Palanker and T Čižmár

Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305, USA
Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
Department of Ophthalmology, Stanford University, Stanford, CA 94305, USA
School of Medicine, University of St Andrews, North Haugh, KY16 9TF, UK

E-mail: ggoetz@stanford.edu

Received 15 May 2013
Accepted for publication 29 August 2013
Published 18 September 2013

Abstract

Objective. We present a holographic near-the-eye display system enabling optical approaches for sight restoration to the blind, such as photovoltaic retinal prosthesis, optogenetic and other photoactivation techniques. We compare it with conventional liquid crystal displays (LCD) or digital light processing (DLP)-based displays in terms of image quality, field of view, optical efficiency and safety.

Approach. We detail the optical configuration of the holographic display system and its characterization using a phase-only spatial light modulator.

Main results. We describe approaches to controlling the zero diffraction order and speckle related issues in holographic display systems and assess the image quality of such systems. We show that holographic techniques offer significant advantages in terms of peak irradiance and power efficiency, and enable designs that are inherently safer than LCD or DLP-based systems. We demonstrate the performance of our holographic display system in the assessment of cortical response to alternating gratings projected onto the retinas of rats.

Significance. We address the issues associated with the design of high brightness, near-the-eye display systems and propose solutions to the efficiency and safety challenges with an optical design which could be miniaturized and mounted onto goggles.

(Some figures may appear in colour only in the online journal)

1. Introduction

Retinal degenerative diseases such as age-related macular degeneration or retinitis pigmentosa are among the leading causes of blindness in the developed world [1, 2]. These diseases lead to a loss of photoreceptors, while the inner retinal neurons survive to a large extent [3–5]. Activation of the remaining neurons can produce visual percepts, known as phosphenes, thereby enabling delivery of information to the visual system in a blind patient. Electrical stimulation of the surviving retinal neurons has been achieved either epiretinally, in which case the primary targets of stimulation are the retinal ganglion cells (RGCs) [6–9], or subretinally to bypass the degenerated photoreceptors and use neurons in the inner nuclear layer (bipolar, amacrine and horizontal cells) as primary targets [10–13].

Recent clinical studies with epiretinal [6] and subretinal [14] prosthetic systems have demonstrated improvements of the visual function in certain tasks, with some patients being able to identify letters with equivalent visual acuity of up to 20/550 [14]. However, both of these systems suffer from a number of shortcomings. Recently approved by the FDA, the ARGUS II implant (Second Sight Inc.) delivers information from an external camera and an external signal processing unit via serial RF telemetry to a sub-conjunctival receiving coil connected to a signal processing module, which then relays the stimulation signals to the epiretinal electrode array via a trans-scleral cable [6]. Such a design is difficult to scale to a much larger number of electrodes than the 60 in the current
model. The bulky receiving coil and processing electronics of the implant make the surgery very difficult, and together with the penetrating intraocular cable make it prone to multiple surgical complications [15]. Furthermore, retinal stimulation in this system is directly determined by images captured by the external camera, disregarding the use of natural eye movements to scan a visual scene, a crucial feature of normal visual perception. Implants with photosensitive pixels, such as that from Retina Implant AG [12], largely overcome the scalability limitation and make use of natural eye movements, but still require external power delivered via RF coils and a trans-scleral cable.

A fully optical design overcomes these limitations by using subretinal stimulating arrays of photodiodes powered by pulsed near infrared (NIR, 880–915 nm) illumination [16, 17]. This approach enables parallel optical transmission of visual information to each pixel in the implant, adjustable stimulation parameters to modulate retinal response [16, 18, 19], and preservation of the natural link between eye movements and visual information. Because the photodiode arrays are operated photovoltaically they do not require any wired power connections, greatly simplifying surgery and reducing complications associated with trans-scleral cables. Each array measures 1.2 × 0.8 mm² in size and 30 μm in thickness, and multiple arrays can be tiled in order to increase the field of view [16]. Other fully optical approaches to restoration of sight include optogenetics, in which retinal neurons are transfected to express light-sensitive Na and Cl channels [20, 21], small-molecule photoswitches which bind to K channels and make them light sensitive [22] or photovoltaic implants based on thin-film polymers [23].

Since all these optical approaches require much brighter illumination than the ambient light can provide, a head-mounted display is required to deliver very bright images to the retina. This system should typically include a camera, a signal processing unit and a near-eye display operating at a wavelength optimized for the particular approach. The camera provides autofocusing and adaptation to a broad range of ambient brightnesses, a feature necessary for all stimulation approaches as they have a much narrower dynamic range than the natural variation of lighting conditions. In addition, this approach provides flexibility for image processing between the camera and the display, which is likely to vary for various target cells and techniques of stimulation, and might require optimization for each patient. For activation of the photovoltaic silicon implants NIR light (880–915 nm) with peak irradiance of several mW mm⁻² [16] is required to elicit retinal response. For other optical approaches very bright UV [22], blue [21], green [23] or yellow [24] light is required.

To enhance portability of the system, reduce its weight and maximize battery life, it is important to minimize optical losses and attain the highest efficiency possible. At the same time, the field of view, resolution, and contrast have to be sufficiently high to create meaningful percepts. Finally, ocular safety concerns ought to be kept in mind when designing a system operating with intense illumination, which is often close to the maximum permissible exposure [16], to ensure that in the event of a critical failure of the display, the user will not be harmed. We describe an optical system based on Fourier holography which provides high power efficiency, very high peak irradiance and allows a sufficiently compact design to be mounted on goggles. We address challenges associated with this geometry, such as the presence of non-controlled light in a zero diffraction order, and complex hologram computation. We show that this system provides good image quality with an acceptable field of view, and demonstrate that image speckling, a very common problem associated with the use of coherent light in imaging, is negligible with our design. We provide a brief safety analysis of the display illustrating the outcomes of potential failure of various components. We provide comparisons with a more conventional LCD-based imaging system throughout this paper. Finally, we demonstrate an application of a slit-lamp mounted version of this holographic projection system through the measurement of cortical response to alternating gratings in normal-sighted rats. Our experiments are tailored to finding an optimal solution for addressing photovoltaic retinal prosthesis [16, 18], but the principles are applicable for optogenetic or other purely optical sight restoration techniques as well.

2. System description

2.1. General principles

Common near-the-eye retinal projection systems (video goggles) are based on modulation of brightness in the image, either by attenuating light intensity in each pixel of the liquid crystal displays (LCD) or by varying the duty cycle of each micro-mirror pixel with digital micro-mirror devices (DMD), as is the case in digital light processing (DLP) technology. Both technologies block and discard all undesired optical power, and within the optical system the irradiance modulating element and the retina are placed in conjugate planes.

The system we designed is a modified version of a standard Fourier holographic geometry based on a nematic phase-only spatial light modulator (SLM), which adjusts the phase of the incident wavefront in each pixel. The phase-modulated optical field is transformed into an image by a lens. Conveniently, the transformation between the modulated optical field and the resulting image can be mathematically described by the Fourier transform [25], which is at the heart of the algorithms for design of holographic modulations. Importantly, instead of blocking light from dark image zones, holographic systems redistribute the optical power across the image with a very high efficiency, thereby greatly reducing power losses in the image formation process, compared to LCD or DMD displays.

A typical layout for a Fourier holographic system is shown in figure 1(a) and corresponds to solutions previously suggested for similar applications, in particular by Golan et al [26]. In such a system, the incident wavefront is modulated by an SLM, and a Fourier lens creates the Fourier transformation of the reflected wavefront in an intermediate image plane. Because of the limited diffraction efficiency of the SLM, undesired diffraction orders, the zeroth order in particular, subsist here. The zero diffraction order affects the central region of the image plane, therefore the image is usually
generated off-center in a spatial region associated with the first diffraction order. This is achieved using a blazed grating, analogous to a prism, superimposed onto the hologram phase modulation, which shifts the first diffraction order to the side of the image. Conveniently, all undesired diffraction orders, including the zeroth order, can be removed by spatial filtering using a physical aperture in the image plane, which blocks all except the first order. The final image is projected onto the focal plane using a telescope. Miniaturizing this classical ‘4f’ Fourier system into a near-to-eye unit is very challenging from an engineering point of view. To reduce the size of the holographic system to video goggles we propose using the configuration shown in figure 1(b). Here, the crystalline lens of the eye is used to Fourier transform the modulated optical field, and the resulting image is projected directly onto the retina. Linearly polarized light is reflected off a polarizing beam-splitter prior to propagating through a quarter wave-plate. It is then reflected off a polarization-insensitive SLM [27], which modulates the wavefront. The quarter wave-plate makes the light linearly polarized again so that it is transmitted by the beam-splitter and enters the eye. A polarization-maintaining fiber allows convenient delivery of optical power through a flexible cable from a light source (laser diode) located along with a battery and a pocket computer in an external unit. The lack of an intermediate image plane in this compact configuration makes it impossible to physically block undesired diffraction orders, therefore the effects of the undesired diffraction orders need to be minimized another way, as described in the next section.

2.2. Defocusing of the zero order

With a well-aligned nematic SLM such as the one used in this study, only the zeroth and first orders carry a significant fraction of the light power, so the effects of the higher diffraction orders can often be ignored for imaging purposes. The situation would be different if a ferroelectric SLM were used, for which there is an equal amount of power sent to the first positive and the first negative diffraction orders, so at the least the zeroth and the first negative order need to be blocked in the intermediary image plane.

With the geometry we propose, there is no intermediary image plane, and therefore blocking these diffraction orders with an aperture is not possible. Whilst careful alignment and proper computation of the hologram can make the higher diffraction orders negligible compared to the first [26], the presence of a zero diffraction order is still unavoidable. With the nematic SLM we used (Boulder Nonlinear Systems, model P512-0785), 77% of the light was diffracted into the first order. Higher efficiency SLMs can increase this fraction to above 90%. Because ferroelectric SLMs divert as much energy into the first positive and negative diffraction orders, they are not well suited to this approach, as over half of the incident light power would be lost.

When the SLM is illuminated by a collimated beam, the zero order appears as a bright focused spot in the image plane. With the power levels required for retinal prosthetic applications, tens of milliwatts incident on the SLM, the presence of a highly focused zero order raises safety concerns. To address this issue, we illuminate the SLM with a non-collimated beam which may be either diverging or converging, and refocus it by adding a suitable quadratic modulation on the SLM, which acts as a Fresnel lens [28]. When the SLM is turned off, all of the light is focused either in front or behind the retina, depending on whether the beam is converging or diverging, as illustrated in figures 2(a) and (b). When the SLM is turned on and a hologram is displayed, the diffracted light is refocused on the retina thus forming an image, while the defocused zero order appears as a uniform background, as illustrated in figures 2(c) and (d). The uncontrolled zero order is still focused either behind (c) or in front of the retina (d), but the diffracted fraction of the light is focused on the target.
image plane. The effect of the zero order background on image quality is discussed below in section 3.

The amount of beam divergence that can be compensated by the SLM is limited by its finite pixel pitch [26]. With a maximum phase shift of $2\pi$, a pixel size of 15 $\mu$m, SLM resolution of $512 \times 512$ pixels and an illumination wavelength of 880 nm, the minimum focal length of the Fresnel lens that can be formed without exceeding the Nyquist sampling frequency in the spatial domain is $\approx 110$ mm. Further increase of the diffraction angle would create aliasing effects producing shifted replicas of the same lens, which would significantly deteriorate the diffraction efficiency and therefore should be avoided. Since this aliasing limitation is determined by the pixel pitch of the SLM, it is common to both concave and convex Fresnel lenses, and thus affects to the same extent convergent and divergent illuminations of the SLM.

2.3. Hologram computation

Algorithms for hologram synthesis have been studied extensively in the past [29]. For our particular application, the algorithm must satisfy two main requirements. First, for testing visual acuity in animals, 1D and 2D grating patterns with large uniform areas will be projected onto the retina, so the algorithm should produce images with good uniformity. Second, holograms should be computed fast enough to provide good substitution for natural vision, with refresh rates likely to exceed 15 Hz.

For our initial studies we selected the Gerchberg–Saxton (GS) algorithm [30], as it is well established, simple to implement, provides good uniformity over the image, and can be computed efficiently using GPU implementations of the fast Fourier transform. We implemented a CPU-based version of the GS algorithm in NI LabVIEW 2012, running on a HP Z420 Workstation. It has been shown that this algorithm allows computation of the holograms at refresh speeds far exceeding video-rates, using a nVidia graphic cards in the OpenGL or CUDA programming environment [31]. With the emergence of the mobile gaming industry, similar capabilities are being developed for mobile devices. Apple Inc. already offers a GPU-accelerated version of the 2D FFT on its iOS mobile operating systems, which would enable real-time computation of holograms on a mobile platform.

3. Performance analysis

3.1. Evaluation of image quality

Lasers are an excellent source of bright illumination, which often result in speckling due to high coherence [32]. LCD and DMD-based systems can use low-coherence, high-power laser diode bars which greatly reduce this phenomenon by coupling multiple mutually-incoherent sources into the same fiber. Unfortunately, this is not an option for SLM-based imaging where high coherence is required. In principle, speckling can be reduced by incorporating a phase-smoothing constraint into the GS algorithm [29]. However, the added computational complexity renders this approach impractical for real-time synthesis of holographic modulations.

Alternatively, speckles can be rapidly shifted and thereby smeared over time by sequentially displaying a series of holograms corresponding to different realizations of the same image. Each hologram of the series can be designed using the GS algorithm with different initial conditions (phase distributions). Rapid alternation of the holograms creating different speckling patterns efficiently blurs the speckles, and the image appears uniform to the observer, but computation of several different holograms for each image increases computational load on the system. Another approach to reduction of speckling is shift-averaging of the holograms [33]. However, all these time-averaging methods can help only in systems having a sufficiently slow response time (e.g. normal photoreceptors, or slow-responding channelrhodopsin in optogenetic restoration of sight) compared to the refresh rate of the display.
Figure 3. Spatial integration of speckles. (a) Photograph of a 70 μm pixel size photovoltaic array showing the three diodes in each pixel. (b) Close-up view of a single pixel. The pixel is composed of three photodiodes in series. (c) Typical checkerboard pattern, each white square is 180 μm in size. (d) Simulated response of the implant with 70 μm hexagonal pixels to the checkerboard pattern shown in (c). Each hexagon corresponds to a pixel in the implant and the color corresponds to the minimal amount of light incident over one of the three photodiodes in the pixel. If one of the photodiodes is shadowed, the pixel appears black, if they are all illuminated, it appears white. (e) Same checkerboard pattern as (c), projected using the SLM system. Speckling makes these dots appear uneven. (f) Response of the device to the checkerboard pattern shown in (e). The response is very similar to the response for the non-speckled checkerboard. (g) Checkerboard with 70 μm size squares. Each square is composed of 6x6 spots. (h) Response to a checkerboard pattern the size of (g). The image has spatial frequencies too high for the device to resolve, resulting in aliasing.

In the case of photovoltaic retinal prosthesis having several diodes connected in series in each pixel [34] the time-averaging approach is not applicable since the response time of photodiodes is orders of magnitude shorter than that of the fastest displays available (a few kHz with ferroelectric SLMs). In these devices, the current generated by each photodiode corresponds to the total light power absorbed. With three photodiodes connected in series in a pixel, the current delivered by each pixel corresponds to the minimum current generated by one of the three diodes: should any one diode in a pixel be shadowed, the entire pixel would be turned off. Since spatial integration instantaneously averages speckling over the photodiode surface it can greatly diminish speckling noise, so long as the speckle size is much smaller than a photodiode. We designed the optical pathway to generate images with speckles two–three times smaller than the width of the photodiodes (corresponding to four–nine speckles per diode area). To assess the effect of speckle integration over diodes of various sizes we acquired images of the projected grating patterns on a CCD camera and simulated the integrated response of the prosthesis to these speckled patterns. The image size was 3 mm × 3 mm, corresponding to the anticipated size of the tiled retinal implants. The results are summarized in figure 3 for an array of 70 μm pixels composed of three photodiodes each. With larger pixels, the speckling noise should decrease in proportion to the diode area.
Figure 3(e) depicts an example of one of the test checkerboard patterns used to evaluate the effect of speckling on image quality. Figure 3(f) shows the computed response of a hexagonal photodiode array to a checkerboard pattern with 170 μm squares. Each hexagonal pixel in this array is composed of three equally-sized photodiodes, as illustrated in figure 3(b). Hexagonal pixels in the array are separated by 5 μm wide trenches. The dark areas are larger than the white areas because when a pixel is on the edge of a checker, it will turn dark as soon as one of its three photodiodes is shadowed. The dominant spatial frequency of the checkers is faithfully rendered, illustrating that speckles integrated over the photodiodes do not interfere with device output. This response can be compared to the response to the non-speckled version of the same checkerboard, as shown in figure 3(d). Qualitatively, both responses are similar, showing that distortions in the response are not due to the presence of speckles but rather to the implant itself. Figure 3(h) depicts the response of the device to the higher frequency pattern shown on figure 3(g). Aliasing clearly appears on this image, demonstrating that the holographic display is able to out-resolve the implant, and that the resolution provided is sufficient to address a photovoltaic retinal prosthesis with 70 μm pixels composed of three photodiodes.

As mentioned above, due to the requirement on compactness and simplicity of the imaging pathway, our geometry does not allow spatial filtering of the unwanted diffraction orders, particularly the zero order, which unavoidably carries a significant fraction of the optical power. Defocusing of the zero order in our system adds a diffuse background illumination over the whole image, thereby decreasing its contrast. Figure 4(a) shows an image with the SLM turned off, where all of the optical power is directed to the zero order. Figure 4(b) shows an SLM generated image (in the first diffraction order) superimposed with the residual zero order.

With a non-zero background illumination, contrast of the image depends on how bright and therefore how sparse the image in the first order is. For grating patterns with 50% white and 50% black areas such as the one in figure 4(b) we obtained contrast of 10:1 within an image area of 1.1 × 1.1 mm². With sparser patterns, white parts become brighter and contrast improves, reaching values in excess of 100:1. Using a higher efficiency SLM would decrease the brightness of the zero order, thereby further increasing the contrast accordingly. For clinical uses of retinal implants, it is likely that information will be displayed using sparse contour images, as the ability of patients to recognize objects in this case is improved [35]. It is therefore realistic to operate under the assumption of a certain degree of sparsity in the image displayed, which helps improve the maximum attainable contrast and has implications on power efficiency and safety of the system (see sections 3.3 and 3.4).

Another potentially detrimental effect of liquid crystal displays is that during transition from one frame to another the state of the display is not well defined, and most of the light is deflected randomly, as illustrated in figure 5. The image in the top left corner corresponds to the initially displayed hologram. Other images in figure 5 illustrate how the light is redistributed as the display transitions from the hologram corresponding to this image to the hologram corresponding to the complementary grating shown in the bottom right corner. The number on each image in the sequence represents the corresponding fractional transition time between the frames. For example, the image in the center with label 50% corresponds to exactly halfway through the transition between the two holograms. During this transition the first grating
Figure 5. Random distribution of light during hologram transitions. The image in the top left corner corresponds to the first hologram, and images in other frames show light distribution during transition of the SLM to the second hologram corresponding to the image in the bottom right corner. The text shows the fraction of the transition time between the two holograms.

image fades, the defocused zero order appears in the center, and then disappears as the new grating pattern emerges.

With continuous illumination, transitions between successive holograms appear as bright flickering flashes of light across the visual field. Rapidly alternating between two different holograms corresponding to the same image makes this flickering appear to the naked eye as a constant background, thereby eliminating a distracting blink between the frames. However, this comes again at the cost of a reduced perceived contrast. Because of the capacitive nature of the coupling between implant and retina, optoelectronic retinal prostheses require pulsed illumination [16, 34], therefore random transitions between the frames should not affect the system performance, as long as pulses of light are applied when the images are properly displayed.

3.2. Field of view

Assuming a $2\pi$ maximum phase modulation per pixel, the addressable area on the image plane is determined by the wavelength and the SLM pixel size. With our $d = 15 \mu m$ pixel pitch and $\lambda = 880$ nm illumination, the maximum deflection angle that the SLM can provide is $\alpha = \pm \tan^{-1}(\lambda/d) = \pm 3.4^\circ$.

In a human eye having a focal length of 17 mm [36], this angular range corresponds to about a $2 \times 2$ mm$^2$ addressable area on the retina, with a corresponding $10^\circ$ diagonal field of view. A smaller pixel pitch would increase the field of view. For example, with a $8 \mu m$ pixel pitch, commercially available on the market today, the maximum deflection angle of $6.3^\circ$ will provide an addressable area of $3.7 \times 3.7$ mm$^2$ on the retina, corresponding to $17^\circ$ diagonal field of view. Since the visual field provided by a SLM-based display depends only on the amount of tilt applied to the incident waveform and the useful fraction of the beam reflected off the display corresponding to the first order can always be kept collimated, increasing the field of view does not change the efficiency of the display. It is important to note that diffraction efficiency can degrade significantly (by up to 60%) at the edges of the addressable square, compared to the center value [27]. Therefore, SLMs with smaller pixel pitch should reduce the areas of reduced efficiency in the corners.

With LCD-based systems the field of view is determined by the size of the display and magnification of the imaging optics used to project it onto the retina. For example, the
Vuzix™ Smart Glasses M100 has 16° diagonal field of view. Having a large field of view and retaining high efficiency is challenging, since very significant clipping of a wide beam occurs on the iris. A Köhler illumination system, in which the illumination beam is focused onto pupil, could be used to minimize the losses.

3.3. Power efficiency of the system

Power efficiency is naturally a very important consideration for a mobile system. Fully optical approaches to restoration of sight to the blind (i.e. without power supply connected to the retinal implant) require high irradiance levels to elicit neural response, which makes the power efficiency even more of a concern.

With a typical dynamic range of an image extending two orders of magnitude (2 log units, or contrast of white to black 100:1), and assuming that, on average, images are in the middle of this range, i.e. having average luminance ten times below the maximum and ten times above the minimum, conventional displays (LCD and DMD) attenuate 90% of the illumination power. SLM-based holographic systems offer an excellent solution to this issue, because unlike LCD or DMD-based approaches, they do not waste optical power by attenuating the illuminating beam in dark areas, but rather redirect it into brighter parts of the image, thereby decreasing power requirements for the illuminating beam by an order of magnitude for a typical image. It is however important to note that for the system to reach its maximum efficiency, a polarization-insensitive SLM is required. Otherwise, a non-polarizing beam-splitter has to be used instead of the polarizing one shown in figure 1, which will lead to a 75% power loss.

The difference in efficiency becomes especially large when sparse patterns are projected, which is likely to be the case in clinical use of a retinal prosthesis [35]. For a pattern which is 1% white and 99% black, the power consumption of the light source used to illuminate the SLM is just 1% of the power required for a LCD system. This difference becomes even larger when sequential activation schemes are used to stimulate the implant. As explained in section 3.4, if for safety reasons each frame displayed by the device needs to be broken into \( k \) subframes, LCD or DMD-based devices will require \( k \) times more energy to activate the implant compared to the single frame activation scheme, because the full beam power has to be used for each of the \( k \) subframes. In comparison, SLM-based devices can use \( 1/k \) of the full power in each of the subframes and the sequential activation scheme does not require more energy than the single frame activation scheme.

Redistribution of the incident optical power into the desired image in SLM systems also helps increase the irradiance of the image to much above the level of the display illumination, unlike LCD or DMD-based systems where the image irradiance cannot exceed irradiance of the display divided by the square of the magnification. For example, assuming equal image size, in a sparse black-and-white image (e.g. contour) having only 1% non-black pixels the illuminated pixels will be 100 times brighter using a SLM compared to an LCD. Brightness control in SLM systems necessitate an active control of the illumination source which changes its power from frame to frame according to the total amount of light in each image. The LCD or DMD systems however typically do not require adjustment of the light source from frame to frame since they attenuate the excessive light intensity in each pixel and white levels stay constant from frame to frame. Such redistribution of the illuminating light allows the use of significantly lower power light sources, which helps not only to improve battery lifetime and compactness of the system, but also to reduce safety concerns regarding absorption of diffused light by the ocular tissues, as discussed below.

Another significant difference between SLM and LCD-based geometries in terms of power efficiency is the fact that with an SLM, the eye is always illuminated with a collimated or nearly collimated beam, regardless of the field of view of the display. Therefore, a beam diameter which matches the pupil size will maximize the transmission efficiency. LCD screens, however, typically use non collimated illumination, which allows viewing of the full width of the display. With a diverging illumination and a field of view of \( \theta \), the maximum fraction of light transmitted through the pupil is proportional to \( 1/\tan \theta \). Thus, increasing the visual field from 10° to 20° will reduce transmission by at least 52%. These power losses can be minimized if a Köhler illumination layout (or other similar design) is used, in which illumination is focused on the iris plane into a spot smaller than the pupil diameter. However, eye movements restrict the applicability of such systems to lateral shifts of an eye in the iris plane that do not exceed the pupil radius.

3.4. Ocular safety considerations

All-optical approaches to restoration of sight to the blind require very bright illumination in order to elicit responses from the target neurons in the retina. While stimulation thresholds for optogenetic [37], photovoltaic polymer [23] or photovoltaic silicon approaches [16] are typically a few orders of magnitude below the maximum permissible exposure for the wavelength used [36], ocular safety limits the range of brightness modulation. For the following discussion we will assume a system with 17° diagonal field of view, corresponding to approximately 3.7 \( \times \) 3.7 mm\(^2\) (with a 5.2 mm diagonal) on the retina for a 17 mm focal length eye. We will consider an example of a 905 nm NIR wavelength, for which the average retinal irradiance limit is 5.3 mW mm\(^{-2}\) [16] (the peak irradiance limit is much higher and depends on pulse duration [16]), but the reasoning can be applied to any other wavelength used for stimulation.

Since with LCD-based display the irradiance in each part of the image can only be attenuated compared to the uniformly illuminated white field, illumination of the LCD screen should be sufficient for supporting delivery of 5.3 mW mm\(^{-2}\) over the whole 5.2 mm diagonal image. The entirety of this power has to fit through the pupil. With invisible near-IR illumination and dark goggles the pupil will likely dilate to at least 3 mm in diameter [36], while the use of bright visible illumination in patients having light sensitivity will likely result in pupil constriction below 2 mm. Therefore, the NIR irradiance in the
pupil plane can reach values in excess of 10 mW mm\(^{-2}\). Even if the beam would fit within the pupil, natural eye movements will result in occasional exposures of the iris to potentially high irradiance levels. To reduce the total power delivered to the eye an eye tracking system could be used, which would turn off the parts of the screen that the user is not looking at [38]. However, if the electronics of the display fail and the whole screen turns white this tracking mechanism cannot prevent exposure of the iris to potentially unsafe irradiance levels. In the case of UV light (used for photoswitches), limitations are likely to apply for the cornea and crystalline lens exposures as well.

For SLM-based displays however, the entirety of the beam power can be concentrated in the image plane in a diffraction-limited spot, thereby providing very high local irradiance with very small beam power. For example, if a certain degree of sparsity in the image displayed is enforced, high irradiances can still be reached on the target image plane while reducing risks of damage on the iris. Should a 90% degree of sparsity be chosen, the total power in the beam required to display the image is then ten times lower than with a LCD or DMD-based display. In the previous example, this would correspond to irradiances on the pupil plane of 1 mW mm\(^{-2}\), compared to 10 mW mm\(^{-2}\) with LCD or DMD displays.

Additionally, sequential activation of the implant can help further reduce the requirements on total beam power. The SLM used in this study has a maximum image refresh rate of 203 Hz, but images will be refreshed much slower on the implant. Assuming a target display refresh rate of 20 Hz, each frame displayed can be divided into up to ten distinct subframes, each used to address 1/10th of the implant. In each subframe, at most 10% of the image is non-zero. With LCD or DMD-based displays, this sequential activation scheme does not translate into a reduction in maximum beam power required. It instead leads to a 10\(\times\) decrease in power efficiency of the system, because the same amount of light has to be delivered to each subframe compared to the single image. However, with a holographic display, each subframe now requires 10\(\times\) less beam power compared to the single frame: no additional power is required to use this sequential activation scheme, and the peak beam power can be reduced by a factor of 20. In the case of a failure of the display control, when the whole display turns white, and assuming the settings described in the previous example, this corresponds to irradiances of 1 mW mm\(^{-2}\) on the pupil plane, compared to 10 mW mm\(^{-2}\) with LCD or DMD displays. Sequential activation of the implant and enforcing sparsity in the desired image can be combined in order to even further reduce requirements on total beam power.

Such use of power is inherently much safer than with the LCD screen-based display, since even with a failure of the screen control the power delivered to the eye remains within safe limits on the iris plane. Because the beam used to illuminate the eye is not collimated, the zeroth diffraction order cannot appear in focus on the retinal pigment epithelium and choroid either, which means that the incident light power is spread over an area corresponding to the full area addressable by the display on these tissues (for example, 3.7 \(\times\) 3.7 mm\(^2\) for an 8 \(\mu\)m pixel pitch SLM), making it inherently safe.

Figure 6. Optical layout of holographic projection system integrated with a slit-lamp. A Laser beam is delivered to projection system via a single-mode fiber (F). The output beam is then partially collimated by a 30 mm achromat doublet (L1), and polarized by polarizer (P1). After reflection on the SLM the image is formed by a \(\sim 100\) mm concave lens (L2) positioned 100 mm away from the SLM. A 60 mm lens (L3) placed 20 mm away from L2 projects the image onto the target plane (I), after reflection on a 50% beam-splitter (BS). The image can be observed through the slit-lamp and the BS. When the SLM is turned off, the slightly diverging beam creates an image of the zeroth order behind the image plane. When the SLM is turned on, the first order appears in focus on the image plane.

4. In vivo testing of the system

4.1. VEP recording setup

As a proof of concept, the holographic display was used to evaluate the visual acuity of Long Evans (LE) rats in vivo. It should be noted that visual acuity is a complex visual function which is dependent on many parameters such as eye refractive power, optical quality of the image on the retina, cortical state, etc. In this paper, we evaluated the visual acuity by patterned visual evoked potentials (VEP) method, similar to that used in babies [39–41].

The projection system shown in figure 6 was mounted on top of a slit-lamp (Zeiss, SL-120). The 638 nm illumination was provided by a 100 mW single-mode laser diode coupled into a single-mode fiber. The polarization of the beam was aligned to the preferred polarization axis of the SLM using a net zero order half-wave plate. Optical distortions in the system, especially those due to the SLM, are compensated using a wavefront correction algorithm [42]. This algorithm can be implemented on the SLM itself and does not require extra hardware. A tandem of lenses (\(f = 100\) mm and \(f = 60\) mm) was used to project the image onto the focal plane of the slit-lamp. The non-polarizing beam-splitter lets the user visually align the system. To allow direct observation of the projected light patterns on the retina via slit-lamp the focusing power of the cornea in the animal eye was cancelled using
a cover slip and viscoelastic material (sodium hyaluronate, ProVisc, Alcon Laboratories).

To record the VEP, three screw electrodes were inserted into the skull and secured with cyanoacrylate glue and dental acrylic, as described in [18]. Two electrodes were placed over the visual cortex of both hemispheres, 4 mm lateral from midline, 6 mm caudal to the bregma. One reference electrode was implanted 2 mm right to the midline and 2 mm anterior to the bregma. Nose and tail needle electrodes served as a reference and the ground, respectively. The signals were recorded using a Diagnosys Espion E3 VEP system, using the built-in pre-amplifiers, amplifiers and line filters. All signals were band-pass filtered between 1.25 and 500 Hz.

Rats were anesthetized with a mixture of Ketamine (75 mg kg$^{-1}$) and Xylazine (5 mg kg$^{-1}$) injected intra-muscularly. Additional injections of 50% of the initial dose were administered every 45 min, or as needed. A heating pad was used to maintain the body temperature at 37.5 ± 0.5 °C. Electrophysiological recordings were conducted with a dim room illumination of 250 nW cm$^{-2}$. All animal care and experiments were carried out in accordance with the ARVO guidelines for the Use of Animals in Ophthalmic and Vision Research and approved by the Stanford Administrative Panel on Laboratory Animal Care. Following pupil dilation and application of the cover slip, the 638 nm laser beam was directed into the eye and the image was visually aligned via slit-lamp, to appear in focus on the retina.

4.2. Effect of hologram swapping on VEP recordings

To assess visual acuity of the animal one-dimensional black-and-white gratings were projected onto the retina, and the square wave contrast was reversed (black-to-white and back) with a frequency of 2 Hz. Amplitude of the steady-state VEP signal at this frequency was used to assess the cortical response. VEP signals decrease with increasing spatial frequency of the patterns, and it reaches the noise level at the limit of visual acuity of the animal.

We first illustrate the effect of the random light redistribution during hologram transitions on VEP recordings. With continuous illumination, this effect resulted in flashes of light at the contrast alternation rate of 2 Hz. This artifact obscured the cortical response to alternations of the stripe positions on the retina. Plots in figure 7(a) show the average of 1000 recordings taken with 2 Hz contrast reversal of two gratings having a very high density of 32 stripes mm$^{-1}$ on the retina, significantly exceeding the visual acuity limit of rats [43, 44]. Despite the fact that rats cannot resolve such a dense pattern, there is a clear response at $t = 170$ ms and $t = 670$ ms, illustrating the effect of flickering caused by the random light redistribution during hologram transitions. The large number of trials required to obtain a signal is explained by the fact that only a small area (1.2 × 1.2 mm$^2$) of the retina was illuminated by the pattern, resulting in a weak VEP signal.

To eliminate this artifact we presented two different holograms for each of the two complementary patterns, alternating at 40 Hz, while the contrast reversal rate remained at 2 Hz, as described in section 3.1. The method described in this section was initially presented to address the speckling problem associated with coherent illumination, but can also be used to reduce the effect of the random light redistribution, as we illustrate here. Flickering at 40 Hz did not induce any noticeable VEP signal, while the contrast reversal at 2 Hz retained a clear response, as illustrated in figure 7(b). VEP recordings for gratings with 32 stripes mm$^{-1}$ on the retina did not produce any signal, while patterns with 8 stripes mm$^{-1}$ show a clear signal at 2 Hz ($t = 160$ ms and $t = 660$ ms). As before, 1000 trials were averaged for this demonstration.

5. Discussion

The holographic goggles system presented in this paper has appealing features compared to more traditional LCD or DMD-based systems for activation of photovoltaic retinal prosthesis or other approaches to restoration of sight that require very bright illumination. We demonstrate that despite the presence of speckles and the zero diffraction order background, it is possible to obtain contrast of 10:1 for images consisting of 50% white and 50% black, and over 100:1 for sparse contour images. The problem of the random light redistribution during hologram transitions can be overcome by high frequency exchange of alternative versions of the holograms encoding the same images. Using this technique, we demonstrated cortical response to motion in rats. However, in applications requiring short-pulse illumination, such as photovoltaic array, proper synchronization of the pulse of light with the display refresh timing will eliminate this problem altogether.

The high brightness requirements for optical approaches to restoration of sight presents at least two significant challenges: ocular safety and battery lifetime, both of which can be addressed using SLM-based display. While LCD or DMD should be illuminated at the maximum expected brightness over the entire area, the SLM can redistribute light into areas as small as a diffraction-limited spot. Therefore, requirements on total beam power can be reduced by orders of magnitude by enforcing a certain degree of sparseness in the patterns projected or using sequential activation schemes, making this approach inherently safer, especially in the event of a failure which would lead to projection of a white screen. Reduction of the average power also translates into a corresponding extension of the battery life.

Holographic displays, however, also suffer from a number of drawbacks. Miniaturizing a traditional Fourier holographic system is challenging from an engineering perspective. The layout presented in this paper is a possible solution to this technical difficulty, but the price paid is in the presence of a zero diffraction order in the final image which cannot be spatially filtered out. However, when a high-efficiency SLM is used, this zero diffraction order does not impact image quality significantly, especially for sparse patterns. Real-time computation of the holograms presents a significant challenge. While current desktop computers are capable of it [31], real-time computation of holograms using the GS algorithm on mobile platforms requires GPU-acceleration which is just becoming available with the iOS mobile platform (Apple Inc.).
Figure 7. Cortical responses (VEP) to optical stimuli projected onto the retina. (a) VEP responses to holograms corresponding to swapping gratings of 32 lines mm$^{-1}$ much higher spatial frequency than the animal can resolve. Clear response at 2 Hz (blue arrows) demonstrates the effect of the light redistribution during hologram transition at this frequency. (b) VEP response to two holograms representing the same image and alternating at 40 Hz, while the grating shift (two other holograms) occurs at 2 Hz. With the grating of 8 lines mm$^{-1}$ the 2 Hz response is apparent (blue arrows). At much higher spatial frequencies (32 lines mm$^{-1}$ on the retina) the signal disappears completely, demonstrating that such spatial frequency is beyond the resolution of the rats eye.

SLM-based systems also suffer from speckling, and the time-averaging methods traditionally used to address this problem do not work in the context of photovoltaic retinal prosthesis due to their very high response speed. Spatial integration of speckles over the size of a photodiode helps reducing this noise, as long as the speckles are significantly smaller than photodiodes. Using the diffraction limit as an estimate of speckle size, speckles in the image plane of our system are about 5–8 μm, corresponding to about six speckles per the smallest diode (20 μm) used in the photovoltaic implants [16]. Direct optical stimulation of single neurons in the retina has much slower response time than photodiodes, and therefore speckles can be averaged over time, using techniques such as shift-averaging [33]. Unlike SLM, the LCD or DMD based systems can use low-coherence illumination sources, such as high-power laser diode bars, resulting in much less-speckled images.
Unlike LCD or DMD-based systems that attenuate illumination in each pixel to form an image, the SLM-based imaging also requires an active control of the light source to adjust the total power for each frame, corresponding to the average brightness of the image.

Several improvements or modifications to this system could be envisioned. Throughout our study, we encoded the Fresnel lens used to collimate the first order over a $2\pi$ modulation. Use of a larger phase shift on each pixel ($4\pi$ or more, available with some SLM models) would allow further defocusing of the zero order and increasing the visual angle: $4\pi$ modulation per pixel would allow twice steeper deflection, which enables collimation of a twice more diverging beam. The zero order could therefore be spread over an area 4 times larger, improving the contrast accordingly. This approach, however, might be accompanied by generation of stronger higher orders along the optical axis, which may interfere with the image.

SLMs also allow correcting systematic aberrations in the optical system, which would otherwise degrade the image quality [42]. This feature could be used to compensate for some aberrations in the patient’s eye, such as astigmatism, near or far-sightedness, and even early stages of cataract. This would require measuring the aberrations of the eye and correcting for them in the computed hologram.

Should hologram computation and speckling become insurmountable obstacles in certain applications, an alternative holographic geometry could be considered, which retains the advantages of high power efficiency and safety at the cost of a more complex optical layout. This geometry is known as the generalized phase contrast method [45] and requires a ‘4f’ configuration. It is a generalization of phase contrast microscopy, in which a Fourier filter is matched to a phase disturbance in order to create high contrast images with high efficiency. In theory, this method does not require computation of holograms, and does not lead to image speckling. However, fabrication of a suitable Fourier filter can be difficult, and it has yet to be demonstrated whether this method can create images with resolution sufficient for a near-to-eye display system.

As a summary, holography is an interesting and promising candidate for solving many of the efficiency and safety concerns that arise with the use of LCD or DMD screens for wearable goggles for prosthetic applications. While the use of a traditional 4f geometry can seem like a difficult engineering challenge at first, it is possible to create holographic display systems based on much simpler layouts. These displays can be made much safer than their LCD or DMD equivalents, and inherently use light in a more efficient way. However, the use of holographic techniques is accompanied by new challenges: speckling and difficult hologram computations. As of today, holographic displays are also much more expensive than their LCD or DMD counterparts, for which the technology is more mature and therefore more affordable. If speckling or cost are a major concern in an application, then currently LCD or DMD-based displays should therefore be preferred, but the resulting system is likely to suffer higher losses, and special attention should be paid to safety concerns due to the dangers associated with Köhler illumination. As the technology develops and matures, it is likely that prices of SLMs will decrease, making the cost issue less pronounced in the future. Future studies will also be dedicated to investigating possible means of minimizing the effects of speckles on image quality.

6. Conclusions

Holographic imaging systems designed for safe and efficient activation of photovoltaic retinal prosthesis enable the projection of contour images with high efficiency, high irradiance and much lower total power than traditional LCD or DMD-based displays. Integration of light over the photosensitive elements reduces speckling noise to acceptable levels for diodes as small as 20 μm. Very compact design of video goggles is based on defocusing of the zero diffraction order, and refocusing the image using Fresnel lens added to the hologram of the encoded image. Solutions to various challenges associated with the holographic approach, such as the presence of multiple diffraction orders, speckles, transitions between the holograms and difficulties in hologram computation were presented. As a proof of concept, the system was successfully tested in vivo by measuring cortical responses to alternating gratings, thus demonstrating feasibility of the holographic approach to near-the-eye display.

Acknowledgments

Funding was provided by the National Institute of Health Grant R01-EY-018608 and by Air Force Office of Scientific Research grant FA9550-04 (DP). TC was supported by the University of St Andrews, UK. TC and GG also received a staff exchange grant from SU2P. We would like to thank H I C Dalgarano for proofreading the manuscript.

References