Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain–machine interfaces

2012 J. Neural Eng. 9 046016

(http://iopscience.iop.org/1741-2552/9/4/046016)

J. Neural Eng. 9 (2012) 046016 (11pp) doi:10.1088/1741-2560/9/4/046016

Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain–machine interfaces

W W Abbott[1] **and A A Faisal**[1,2]

[1] Department of Bioengineering, Imperial College London, SW7 2AZ London, UK
[2] Department of Computing, Imperial College London, SW7 2AZ London, UK

E-mail: aldo.faisal@imperial.ac.uk

Received 14 February 2012
Accepted for publication 8 June 2012
Published 12 July 2012
Online at stacks.iop.org/JNE/9/046016

Abstract

Eye movements are highly correlated with motor intentions and are often retained by patients with serious motor deficiencies. Despite this, eye tracking is not widely used as control interface for movement in impaired patients due to poor signal interpretation and lack of control flexibility. We propose that tracking the gaze position in 3D rather than 2D provides a considerably richer signal for human machine interfaces by allowing direct interaction with the environment rather than via computer displays. We demonstrate here that by using mass-produced video-game hardware, it is possible to produce an ultra-low-cost binocular eye-tracker with comparable performance to commercial systems, yet 800 times cheaper. Our head-mounted system has 30 USD material costs and operates at over 120 Hz sampling rate with a 0.5–1 degree of visual angle resolution. We perform 2D and 3D gaze estimation, controlling a real-time volumetric cursor essential for driving complex user interfaces. Our approach yields an information throughput of 43 bits s^{-1}, more than ten times that of invasive and semi-invasive brain–machine interfaces (BMIs) that are vastly more expensive. Unlike many BMIs our system yields effective real-time closed loop control of devices (10 ms latency), after just ten minutes of training, which we demonstrate through a novel BMI benchmark—the control of the video arcade game ‘Pong’.

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

1. Introduction

The advancement of brain–machine interface (BMI) technology for controlling neuromotor prosthetic devices holds the hope to restore vital degrees of independence to patients with neurological and motor disorders, improving their quality of life. Unfortunately, emerging rehabilitative methods come at considerable clinical and post-clinical operational cost, beyond the means of the majority of patients [1]. Here we present an ultra-low-cost alternative using eye-tracking. Monitoring eye movement provides a feasible alternative to traditional BMIs because the ocular-motor system is effectively spared from degradation in a wide variety of potential users, including those with muscular dystrophies and motor neuron disease [2, 3]; spinal traumas, because ocular innervation comes from the brain-stem; paralysis and stroke, when brain lesions occur in areas unrelated to eye movements; amputees; multiple sclerosis and Parkinson’s, which affect eye movements later than the upper extremities; as well as for a rapidly ageing population with longer life-spans that usually results in progressive deterioration of the musculoskeletal system. The ability to control eye-movements can therefore be retained in cases of severe traumas or pathologies in which all other motor functions are lost. Based on the disease statistics, we find that within the EU alone, there were over 16 million people in 2005 (3.2% of the population) with disabilities who would benefit from such gaze-based communication and control systems [4].

We observe the world through discrete, rapid, focussed eye movements (saccades) acting to align the high resolution central vision area (fovea) of both eyes with an object of interest (fixation point). Visual information is vital to motor planning and thus monitoring eye-movements gives significant insight into our motor intentions, providing a high frequency signal directly relevant for neuroprosthetic control. Eye tracking and gaze-based human–computer interaction is a long established field; however cost, accuracy and inadequacies of current user interfaces (UI) limit them to their more common use in clinical diagnostics and research settings. Low cost eye-tracking systems have been developed by others using off the shelf web-cams [5–7]. However, the performance of these systems still does not match commercial grade systems. This is due to different combinations of low frame rate ($\leqslant$30 Hz), resulting in motion blur and missing saccades—requiring sample frequencies >100 Hz; and poor gaze angle accuracy and precision, leading to an unreliable, noisy gaze estimate. Currently, high performance commercial eye-tracking devices (system cost >20 000 USD) are primarily used to record eye-movement for academic or industrial research. This is because in addition to cost, there are remaining issues surrounding the effective integration of eye tracking into gaze-based interaction systems for everyday patient use. Fundamentally, gaze-based interaction requires the differentiation of normal behavioural eye movements and intentional eye 'commands', which is known as the Midas touch problem [8]. This is a major issue for existing gaze-based computer interaction, which focus on monocular eye tracking to drive a mouse pointer. The 'select or click' command is usually derived from either blink detection or gaze-dwell time, both of which also occur in natural behaviour and thus require an extended integration time (typically in the order of seconds) to initiate a reliable click. We have developed an ultra-low-cost binocular eye tracking system that has a similar accuracy to commercial systems and a frame rate of 120 Hz, sufficient to resolve saccadic eye movements (frequency ~100 Hz). We have addressed the Midas touch problem by distinguishing non-behavioural eye winks from behavioural eye blinks, significantly speeding up selection time. In the future we aim to use eye movements to control motor prosthesis for restoring independence to severely disabled patients. The major challenge here is to derive a practical control signal from eye movements that meets the interface requirements. This must be achieved without being intrusive to the natural sensory function of the eyes. We aim to allow the user to interact with their surroundings directly rather than limiting their interactions to via their computer visual display unit (VDU). Towards this, we derive a BMI signal that provides an information rich signal for inferring user intentions in natural contexts: three dimensional (3D) gaze position.

We interact with a 3D world, navigating and manipulating our surroundings. Severe disabilities remove this ability, vital for independence. Gaze-based interaction for computer control works towards restoring this by facilitating interaction with the world via a computer VDU; instead we propose direct 3D gaze interaction for motor-prosthetic control. With knowledge of both eye positions, gaze-depth information can be obtained because the eye vergence system forces both eyes to fixate on the same object, allowing image fusion and depth perception. The intention-relevant, high-information throughput 3D gaze signal can be applied to tasks such as wheelchair navigation, environmental control, and even the control of a prosthetic arm. Despite the huge potential, 3D gaze estimation has received less attention than the 2D alternative for (mouse) cursor control. A major challenge of gaze estimation, particularly in 3D, is the calibration and adaptation of the estimation system for individual users. Existing 3D approaches can be divided into virtual and non-virtual methods. Interaction with 3D stereoscopic displays using gaze estimation to make icon selection in the virtual volume has received some attention. For these virtual applications, there are currently two main calibration approaches: (1) calculating the intersection point between the monocular gaze vector and the known virtual 3D surfaces [9] and (2) obtaining 3D calibration points to learn a mapping between binocular eye positions and a virtual 3D gaze location [10]. These methods can only be applied with a 3D stereoscopic display.

Gaze interaction with the non-virtual 3D environment has received less attention, though Hennessey and Lawrence in 2009 developed the first binocular gaze tracking system for estimating the absolute *X*, *Y*, *Z* coordinates of gaze targets in the real 3D world [11]. Their method uses the explicit geometry of the eye and camera mounting to relate the pupil position in each camera image to the 3D gaze vector of each eye. The gaze vector is the ray that runs between the centre of the fovea in the retina, through the cornea to a gaze fixation point (neglecting the kappa offset between the visual and optical axis). To obtain the gaze vectors requires precise positioning of the cameras with full geometric parameterization of the hardware setup; optical properties and a model of the eye, including the refractive index of the fluid inside the eyeball (vitreous fluid). Based on the vergence system, a 3D gaze fixation point is then calculated from the gaze vectors' nearest point of approach. This system has only been demonstrated in controlled research environments, possibly because of the strict geometric requirements and detailed modelling of the physical system.

These existing methods (virtual and non-virtual) are suitable for the controlled settings of their proposed applications, but limit their practicality for motor prosthetic interfaces. Stereoscopic displays are expensive and not very portable, while the precision setup of geometric methods is not feasible with low cost hardware. We present here our portable ultra-low-cost hardware with a suite of algorithms and realization of a system that can estimate the absolute gaze target in *X*, *Y* and *Z* coordinates with an accuracy that rivals present methods, without complex configuration routines, the need for 3D display equipment or user-specific details about eye geometry.

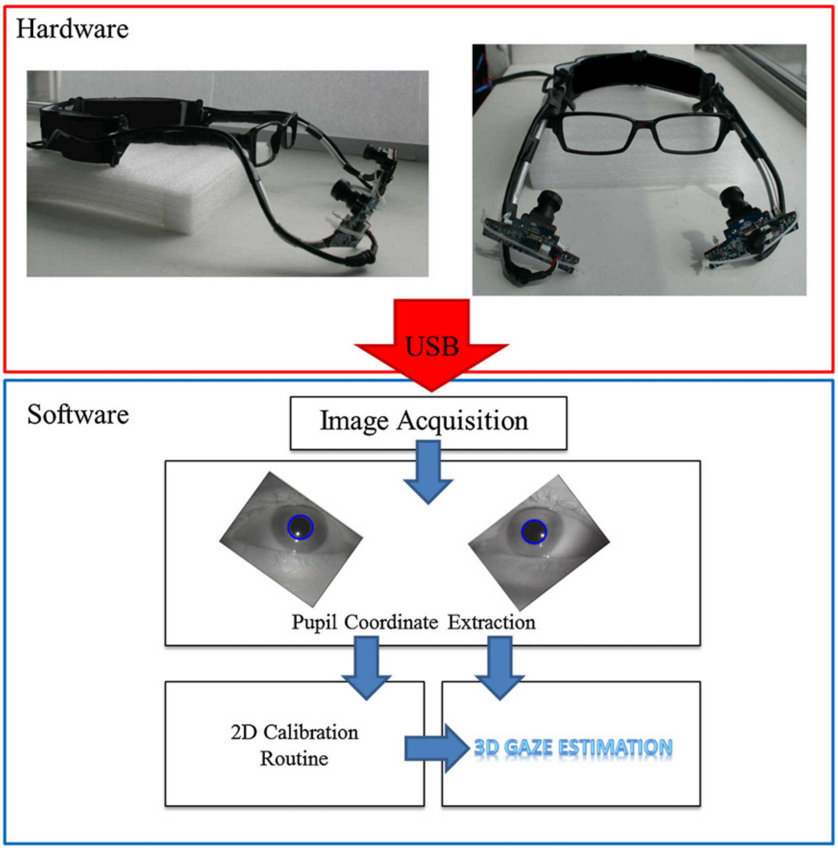

Figure 1. System overview. Hardware: ultra-low-cost head mounted binocular eye tracker built using off the shelf components including two PlayStation 3 Eye cameras (10 USD each), two IR LEDs, cheap reading glasses frames and elastic headband support. The cameras are mounted on lightweight aluminium tubing. The hardware total cost is 30 USD. Software: the camera frames are streamed at 120 Hz via USB to a standard laptop computer and the pupil positions are extracted using image processing (see figure 3). A 2D user calibration allows a mapping between pupil and 2D gaze position to be learnt. Using the 2D estimates from both eyes, a 3D gaze estimation can be made by estimating the vergence point.

2. Methods

Our presented system is composed of ultra-low-cost imaging hardware and stand-alone software that implements our algorithms and methods for 3D and 2D gaze tracking.

2.1. Ultra-low-cost binocular eye-tracking hardware

The video-based binocular eye tracker shown in figure 1 uses two ultra-low-cost video game console cameras (PlayStation 3 Eye Camera—10 USD per unit), capable of 120 Hz frame-rate, at a resolution of 320 × 240 pixels. This is the main cost-reducing step in our system, as typical machine vision cameras operating at this performance are more expensive by two orders of magnitude. To optimize imaging conditions, we modified the camera optics for infrared (IR) imaging at no material cost by removing the IR filter and replacing it with a piece of exposed and developed film negative which acts as a low-cost IR-pass filter. We illuminate the eyes using two IR LEDs aligned off axis to the camera, creating a dark pupil effect to enhance the contrast between the pupil and the iris. Chronic IR exposure above a certain threshold leads to retinal damage or the formation of cataracts [12]. This threshold has been reported as being between 10 and 20 mW cm^{-2} [12, 13]. The LEDs used are Optek gallium arsenide OP165D which produce an irradiance of between 0.28 and 1.6 mW cm^{-2} (depending on the forward voltage) measured at a distance of 1.5 cm. This is well below the safety parameter, especially as it will be mounted at 10 cm from the eye. These LEDs are powered using a USB cable giving a 5 V supply with up to 500 mA of current to be drawn. The driver circuit provides 20 mA current to each LED with a forward voltage of 1.6 V applied. The cameras are head mounted to maximize the eye image resolution and allow unrestrained head movement following calibration. The camera-mounting headset shown in figure 1 has been designed with off the shelf components costing 10 USD in total. The system weighs in total 135 g and the cameras and their mounting arms exert a moment of approximately 0.1 Nm on the nose. It has been designed to allow four degrees of freedom for adjustment to different users (shown in figure 2). The images from the cameras are streamed via two USB 2.0 interfaces to a standard laptop computer facilitating an accessible and portable system.

2.2. Eye tracking

The eye-tracking methodology applies standard image-processing methods to locate the pupil centre in each video frame; an overview of this process can be seen in figure 3. The IR imaging system increases the contrast between the

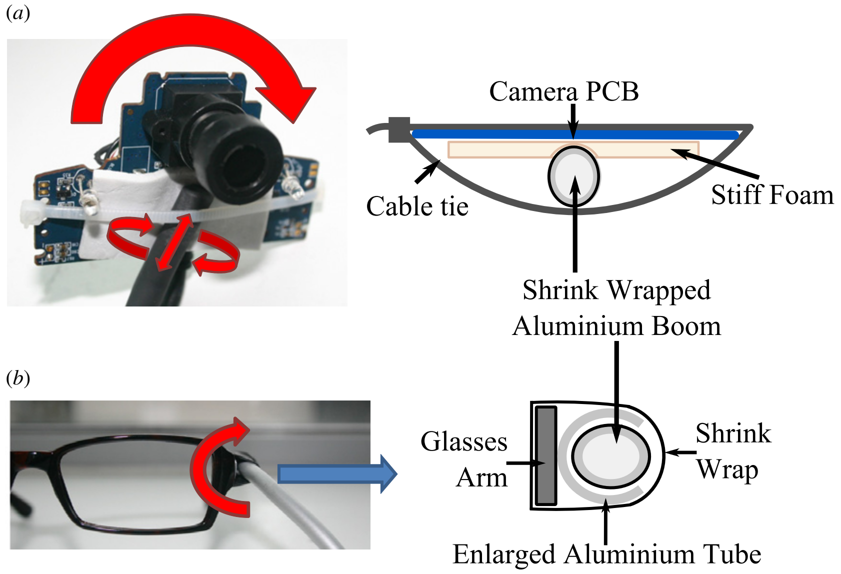

Figure 2. Headset adjustability. Headset design allows the camera position to be adjusted with four degrees of freedom: (*a*) rotation and translation of the camera on the boom arm and (*b*) rotation of the boom arm itself. This allows adjustment to customize the system to different users.

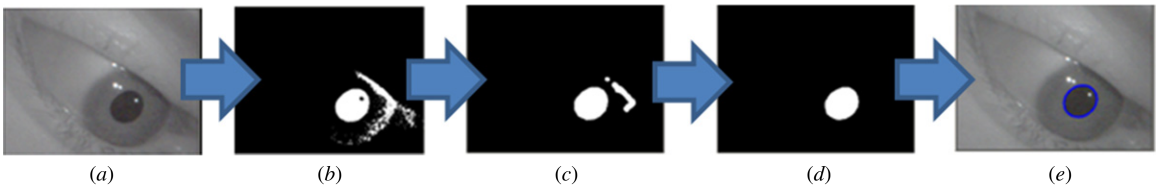

Figure 3. Pupil extraction image processing pipe-line. Images intermediates include: (*a*) raw greyscale, (*b*) binary, (*c*) noise filtered, (*d*) shape filtered, (*e*) original with extracted ellipse overlaid.

pupil and iris. This allows simple intensity threshold image segmentation, converting the greyscale image to a binary image (figure 3(*b*)). In this single step, the data volume per frame is reduced from 230 to 9.6 kB; retaining sufficient information to locate the pupil but reducing the subsequent computational load. Due to noise effects and other dark regions in the image, such as shadows and eyelashes, a pupil classification step is made. To reduce the complexity of the classification process, morphological operations of erosion and dilation were applied in a sequence—first 'opening' the image, removing the dropout noise; and then 'closing' it to fill in any holes in the pupil blob (figure 3(*c*)). Connected component labelling is then applied to assign a unique label to the pixels of each candidate pupil region. Subsequently, a shape-based filter is applied (figure 3(*d*)) to classify the pupil-based on maximum and minimum object size and elongation (axis ratio). The pupil centre is then extracted using least-squares regression to fit an ellipse to the classified pupil object contour—this ellipse is shown overlaid on the raw image in figure 3(*e*). The x and y coordinates of the ellipse centre (in pixels) are extracted for each eye ellipse as the pupil positions.

2.3. Calibration for 3D gaze estimation

The pupil positions extracted from the eye images must be related to the gaze position. A purely explicit method requires a rigid system setup difficult to obtain using low cost hardware, while a purely implicit method requires more involved 3D calibration points. We achieve gaze estimation in the real 3D environment by combining an implicit step to infer the system parameters with an explicit geometric step to transfer this to a 3D gaze estimate. This involves the calibration of each pupil position to the respective gaze positions on a computer VDU (2D calibration). From this, the 3D gaze vector of each eye can be found (step 1) from which the 3D fixation point is then calculated (step 2).

Step 1: Calculating 3D gaze vectors using 2D calibration. Calibration to the 2D computer monitor can be made explicitly using the geometry of the system [14–17] or implicitly using a calibration routine to infer a mapping between pupil position (in the eye image) and gaze position (on the computer screen) [18–20]. The implicit mapping provides a more suitable solution because explicit methods require precise geometric knowledge of camera positions, infeasible with the low-cost adjustable headset. To learn an implicit mapping, training data is acquired using a calibration routine which displays each point of a 5 × 5 calibration grid that spans the computer VDU. At each calibration point, the pupil location of each eye is extracted from a ten-frame burst and the user's average eye positions are recorded. This reduces noise effects of drift and micro-saccades. The calibration data points collected are used to train a Bayesian linear combination of nonlinear basis functions. The second-order polynomial basis functions were found to achieve an optimum trade-off between model

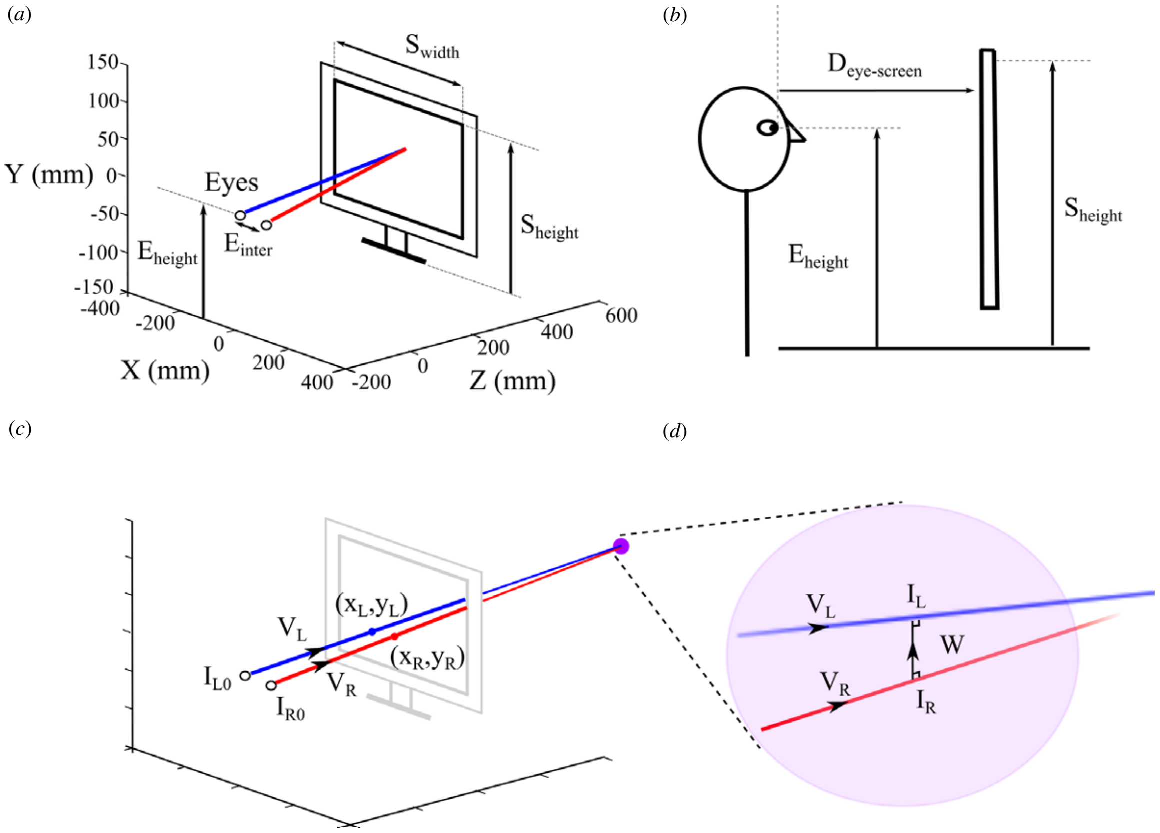

Figure 4. Illustration of our 3D gaze estimation method. (*a*) 2D calibration step to relate the pupil positions to their gaze positions in the VDU screen plane. The user is aligned with the horizontal screen centre and must remain stationary during calibration. The measurements shown are required for calibration. (*b*) Side view of user and computer VDU screen. (*c*) The left and right gaze estimates on the VDU are represented by the two dots (x_L, y_L) and (x_R, y_R) and yield the gaze vectors shown (V_L and V_R). (*d*) The nearest point of approach on each gaze vector is found.

complexity and the number of calibration points required to generalize well. Following the 2D calibration routine, the second-order polynomial mapping is used to map the position of each eye to the gaze position in the 2D plane of the computer monitor, at a frame rate of 120 Hz. When the user fixates in the monitor plane, the gaze estimates of each eye are approximately superimposed as shown in figure 4(*a*). When the user fixates outside of the monitor plane, the 2D gaze estimates diverge as shown in figure 4(*c*). This divergence gives depth information as the 2D gaze estimates are effectively the intersection between the gaze vectors and the computer monitor plane (see figure 4(*c*)). The gaze vectors are calculated from the 2D gaze estimates x_L, y_L and x_R, y_R (relative to the top left corner of the screen) using equations (2.1) and (2.2). This requires the relative positions of the eyes and monitor to be fixed during calibration and measurements of screen height (S_{height}), eye level (E_{height}), eye to screen ($D_{\text{eye-screen}}$) and inter-eye distance (E_{inter}) to be made (see figures 4(*a*) and (*b*)). The eye tracker is head mounted and the system is calibrated with a head-centric coordinate system; thus, following calibration the user will be free to move his/her head and the gaze vectors will be relative to the origin which lies between the eyes.

Step 2: Using the 3D gaze vectors to estimate the 3D gaze position. We use the 3D gaze vectors to estimate the 3D gaze position. The 3D gaze position is the vergence point of the two gaze vectors. Exact 3D vector intersection is unlikely; thus, the nearest point of approach on each vector is found. These points are represented by I_L and I_R in figure 4(*d*) and are given by the parametric equations (2.3) and (2.4). The positions of the eyes relative to the origin are represented by I_{L0} and I_{R0} as shown in figure 4(*c*) while S_L and S_R represent scalars to be found:

$$\vec{V}_L = \begin{bmatrix} x_L - \dfrac{S_{\text{width}}}{2} + \dfrac{E_{\text{inter}}}{2} \\ S_{\text{height}} - E_{\text{height}} - y_L \\ D_{\text{eye-screen}} \end{bmatrix} \tag{2.1}$$

$$\vec{V}_R = \begin{bmatrix} x_R - \dfrac{S_{\text{width}}}{2} - \dfrac{E_{\text{inter}}}{2} \\ S_{\text{height}} - E_{\text{height}} - y_R \\ D_{\text{eye-screen}} \end{bmatrix} \tag{2.2}$$

$$I_L(S_L) = I_{L0} + S_L\vec{V}_L \tag{2.3}$$

$$I_R(S_R) = I_{R0} + S_R\vec{V}_R. \tag{2.4}$$

By definition, the nearest points of approach will be connected by a vector that is uniquely perpendicular to both gaze vectors—$\vec{W}$ shown in figure 4(*d*) and equation (2.5). To satisfy this condition, the simultaneous equations (2.5)–(2.7) must hold. Substituting equations (2.3) and (2.4) into the simultaneous equations and solving for the scalars S_L and S_R, we can then obtain the nearest points of approach—I_L and I_R.

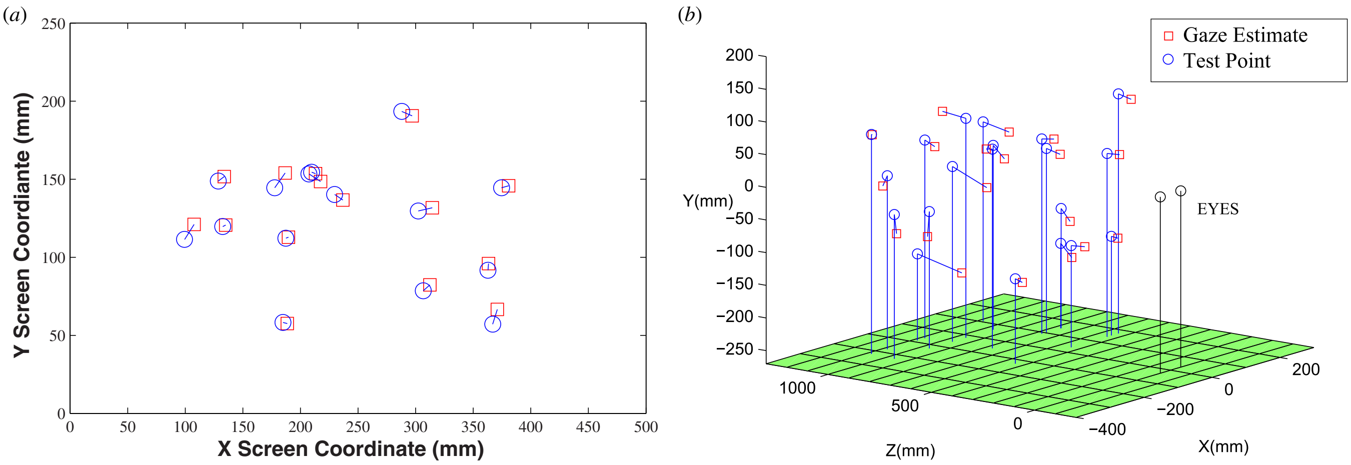

Figure 5. 2D and 3D gaze estimation test points and gaze estimates. (*a*) 2D gaze estimation: the circles represent the 15 randomly generated test positions displayed on the VDU and the squares are the gaze estimate. (*b*) 3D plot of 3D gaze estimation results for a calibration test run with test points being displayed at four depths—(54, 77, 96 and 108 cm) following 2D calibration at 54 cm. The circles represent the real displayed positions and the squares represent the gaze estimations.

The 3D gaze estimation (G_{3D}) is then taken as the mid-point of these two positions as shown in equation (2.8):

$$\vec{W} = I_L - I_R \tag{2.5}$$

$$\vec{W} \cdot \vec{V}_L = 0 \tag{2.6}$$

$$\vec{W} \cdot \vec{V}_R = 0 \tag{2.7}$$

$$G_{3D} = \frac{I_L(S_L) + I_R(S_R)}{2}. \tag{2.8}$$

This algorithm is performed for each frame in the video streams to obtain a 3D gaze estimate at 120 Hz sampling frequency that can be used as a volumetric cursor in the development of advanced user interfaces for neuromotor prosthetics.

2.4. Solution to the Midas touch problem

To address the Midas touch problem, the system uses non-behavioural winks to confirm gaze commands. Winks can be distinguished from behavioural blinks by virtue of the binocular eye-tracking feed allowing for much shorter command integration times. In the filtered binary eye image (see figure 3(*d*)), when the eye is closed, no pupil object is located by the eye tracker, raising a closed eye flag. We distinguish between a left eye wink, right eye wink and simultaneous eye blinks using temporal logic. For example a left wink is defined by the left eye being closed and the right eye being open simultaneously for more than 20 frames. The high frame rate allows for this distinction to be made reliably with low integration times of ∼170 ms.

3. Results

We group the three main contributions of this paper into the following sections. (1) High performance ultra-low-cost binocular eye-tracking system: *eye-tracking accuracy and precision*. (2) Solution to the Midas touch problem and continuous control: *human computer interaction and a BMI benchmark for closed-loop control of devices.* (3) Gaze estimation in the 3D environment: *accuracy and precision in 3D tasks.*

3.1. Eye-tracking accuracy and precision

To precisely estimate the eye tracking system's accuracy, a subject was calibrated and shown random test points of known 3D locations in three separate trials. Each trial involved a calibration routine that cycled through a 5 × 5 calibration grid displayed on a computer monitor 50 cm from the user's eyes, followed by 15 randomly generated test points. The results for one trial are shown in figure 5(*a*). For each trial, a new set of random test points was generated. Each test point appeared in turn and the user looked at the point and hit the space bar, at which point the gaze position was recorded from the real-time data stream. Over all trials a mean Euclidean error of 0.51 ± 0.41 cm (standard deviation) was achieved at a distance of 50 cm which translates into an angular error of 0.58 ± 0.47 degrees.

Table 1 compares our system with a commercially available binocular eye tracking system. Our gaze angle accuracy was 0.58 deg ± 0.47 (mean ± SD) and is defined as the eye-tracking signal accuracy, namely how precisely the viewing direction of the eye can be determined. This measure is viewing distance and application-independent but has direct implications for both 2D (monocular) and 3D (binocular) gaze target position estimation accuracy. We achieved an average of 0.58° while the reference system's EyeLink II manufacturer specifies a typical average accuracy as <0.5°—but do not provide more specific data or measurement approach. Our system can perform 2D or 3D gaze estimation, while the EyeLink II, though also a binocular eye-tracker, has software to perform 2D estimation only. Our system is less than one-third of the mass at 135 g compared to the 420 g commercial system, and less than 1/800th of the cost with a unit cost of just 30 USD compared to the 25 000 USD commercial system

Table 1. Comparison between our system (referred to here as GT3D) and the commercial EyeLink II. Here we make comparisons using the metrics in the EyeLink II technical specifications. More detailed analysis of the 3D performance is shown in table 3.

Metric	GT3D	EyeLink II[a]
Gaze angle accuracy	0.58 ± 0.47°	<0.5° [b]
Gaze estimation modes	2D, 3D	2D
Horizontal range	34°	40°
Vertical	20°	36°
Headset mass	135 g	420 g
Frame rate	120 Hz	250 Hz
Cost	30 USD	25 000 USD

[a] Information is taken from the SR Research issued technical specification.

[b] For the EyeLink II the accuracy is expressed as a 'typical average' in corneal reflection mode. No error measurement is given. We provide here the mean gaze angle accuracy and standard deviation for our GT3D system averaged over three separate trials.

cost. Since the tracking range is less for our system—6 degree smaller horizontally and 16 degree smaller vertically—it is an image-processing problem that needs to be solved. The frame rate is also slightly lower at 120 Hz compared to the 250 Hz of the commercial system, but is sufficiently high to resolve saccades and gives a frame rate four times that of other low-cost systems.

3.2. Human computer interaction and a BMI benchmark for closed-loop control of devices

With the system in 2D mode, the user can operate a computer, performing such tasks as opening and browsing the web and even playing real-time games. The system does not require a bespoke graphical user interface, operating in a windows environment with an icon size of 3 cm^2. The use of wink commands allows the integration time to be reduced to ~170 ms of wink to make a selection. To demonstrate real-time continuous control using the eye-tracking system, we used it to play classic video game 'Pong'. This is very simplified computer tennis where the user has to return the ball by moving a racket to meet the approaching ball. We chose this very simple game because it can be used with a mouse input, played against a computer opponent and is readily available online. This allows it to be used as a very simple benchmark that other BMIs can be tested against.

We conducted a user study (six subjects, aged 22–30) to test closed-loop real-time control performance for our interface. Subjects (five first-time eye-tracking users) were calibrated using our 5 × 5 grid method and then given 10 min to learn to play Pong and to get used to using their eyes as a control input. Subject hands asked to keep their hands folded in their laps. The tracked gaze position (vertical component) controlled the Pong paddle position on the screen, as the paddle followed the movements of the mouse pointer (which we directly controlled through GT3D). Thereafter subjects played four full games of Pong against the computer (up to a score of nine points) using our interface and then four full games using their hands to control the computer mouse. The final score and number of returned shots are reported in table 2. On average subjects using our gaze-based approach achieved a score of 6.6 ± 2.0 (mean ± SD across subjects) compared to the computer opponent score of 8.4 ± 1.30 (mean ± SD across subjects). Subjects made 43 ± 16 successful returns per game and on average 25% ± 14% of games were won by the player, i.e. at least one game won out of four. Using the mouse input, subjects achieved a mean score of 8.3 ± 1.5 against computer opponent 5.5 ± 2.7, with a mean of 53 ± 14 player returned shots and 80 ± 22% of games won by the player. In addition, we compared the zero-control scores (without any user input) giving an average score of 0.5 ± 1.0 and computer opponent score of 9.0 ± 0.0 with a mean of 8.0 ± 3.6 returned shots per game.

Table 2. Pong gaming performance. For each input modality the mean and standard deviation for the player and computer scores, number of returned shots per game and percentage of player wins.

	Score			
	Player	Computer	Returned shots	Total player wins (%)
Our system	6.6 ± 2.0	8.4 ± 1.3	43 ± 16	25 ± 14
Mouse input	8.3 ± 1.5	5.5 ± 2.7	53 ± 14	80 ± 22
No input	0.50 ± 1.0	9.0 ± 0.0	8.0 ± 3.6	0 ± 0

Table 3. Three-dimensional gaze estimation performance for the results shown in figure 5.

	Mean absolute error (cm)	Standard deviation (cm)
x	1.1	0.7
y	1.2	1.1
z	5.1	5.0
Euclidean	5.8	4.7

These scores form the framework for our proposed new benchmark of closed-loop real-time control for BMIs and we make the ready-to-use browser-based game available through our website to facilitate benchmarking BMI systems (http://www.FaisalLab.com/Pong). As well as the above scoring metric, the benchmark participants should include the amount of prior training and acclimatization time of the BMI system (for our system this is 10 min) and system/treatment costs.

3.3. Accuracy and precision in 3D tasks

To assess the 3D gaze estimation, the methodology was similar to the 2D experiment but test points were also displayed at different depths. Following the 5 × 5 2D calibration routine at a depth of 54 cm, five random test points were generated at four depths: 54, 77, 96 and 108 cm by moving the computer monitor. Over this workspace, the system performed with a mean Euclidian error of 5.8 cm, with a standard deviation of 4.7 cm. The results for this experiment are displayed on a 3D plot in figure 5(*b*). The mean absolute error and standard deviation for each dimension is shown in table 3. The mean depth error (Z in table 3 and figure 5(*b*)) is 5.1 cm with a standard deviation of 4.7 cm; this accuracy and precision is four times larger than the horizontal and vertical

Table 4. Our 3D gaze estimation performance (referred to here as GT3D) comparison with Hennessey and Lawrence's system [11]. Mean Euclidian errors and standard deviation in cm and as a percentage of the workspace depth[a].

	Mean Euclidean error (cm)	Mean Euclidean error (%)[a]
GT3D	5.8 ± 4.7	5.3 ± 4.4
Hennessey and Lawrence	3.9 ± 2.8	9.3 ± 6.7

[a] Measurements normalized to the workspace depth over which the methods were tested. GT3D—108 cm from the user. Hennessey and Lawrence—the workspace is described relative to the corner of their computer screen rather than the user's eyes. To allow comparison between our data and their data, we assume the subject's head was 60 cm away from the screen and we know that the depth closest to the screen is 17.5 cm away which yields a workspace depth of 42.5 cm = 60–17.5 cm.

equivalent (X and Y in table 2 and figure 5) which explain the considerably higher Euclidean error in 3D compared to 2D gaze estimation. The gaze angle fluctuates around a value of $0.8 \pm 0.2^\circ$ (mean ± standard deviation) but does not consistently increase with depth. The mean depth error for estimations at each depth increased from 4.6 cm at 54 cm distance from the face to 6 cm at 108 cm distance. This is to be expected as a consistent gaze angle error will cause a larger spatial error at deeper depths, particularly in the depth direction.

4. Discussion

We have developed the first ultra-low cost, high-speed binocular eye tracking system capable of 2D and 3D gaze estimation, costing 1/800th of a reference commercial system that achieves a comparable eye tracking performance. This system drives a mouse-replacement based user-interface for which we have implemented an improved solution to the 'Midas touch problem'. Tracking both eyes allows for 'wink' rather than 'blink' detection, decreasing the required selection integration times by a factor of 6. This is because when blink or dwell time is used to make a selection, we must blink or dwell for an extended period of time to distinguish commands from normal behavioural blinks and fixations. The system interfaces with the computer operating system via USB, and allows the user to browse the web, type on a visual keyboard and play real-time games. We demonstrate closed-loop performance by playing a version of the 2D video game Pong (see below).

In the 3D domain, gaze estimation is directly applicable to motor prosthetics, with the potential to allow patients to interact with their surroundings. Our system can estimate the absolute real-world 3D gaze position in real time with a performance competitive with research systems. Table 4 shows the performance comparison of our system with Hennessey and Lawrence's system [11]. As the table shows, the mean Euclidean error of our system is almost 2 cm higher. Hennessey and Lawrence calibrate their system using calibration points taken at both the nearest depth (17.5 cm) and the farthest (42.5 cm). While for our system we calibrate at a single depth of 54 cm and our work space extends out to 108 cm depth from the eyes. The workspace used by Hennessey and Lawrence covers half the depth (25 cm compared to 54 cm) and though Hennessey and Lawrence do not give the workspace depth from the face explicitly (their coordinate system has its origin at a corner of the computer monitor) we estimate the maximum workspace depth to be 42.5 cm from the eyes (see table 4 footnote) compared to the 108 cm depth we used. At larger depths we expect the estimation accuracy to be poorer, as such we make a more balanced comparison by normalizing the error and standard deviation by the workspace depth as can be seen in the second column of table 4. With this metric we see that our system performs with almost half the normalized Euclidian error of their system. For both methods, we see that the error has a large standard deviation, with a magnitude similar to the mean. This variability is partly due to noise in the image sensors and head-set slippage, and may also be due to micro-saccades and drift movements of the eyes.

We found that the gaze estimation error increased linearly with gaze target depth (the way we expected for our method) as determining the gaze intersection point from both eyes would be limited by gaze angle accuracy. This relationship also held for the system presented by Hennessey and Lawrence [11], except for the test depth closest to the face, for which they reported an increase in error. While they assumed the gaze angle accuracy to be constant across depths, we measured and found our system's gaze angle accuracy to be uncorrelated with depth. In the future, it will be important to measure and compare eye-tracking systems used in BMI contexts in terms of their calibration strategy and the effect of behavioural and anatomical differences between subjects (e.g. [11] pooled data across seven subjects). A systematic large user group study, beyond the scope of this proof-of-principle paper, will enable us in future to extract priors for the natural statistics of gaze target to enable applying empirical data for principled Bayesian gaze target estimation.

Although the performance is similar, we require only a standard computer monitor as opposed to 3D equipment, no information on eye geometry, optics or precise camera positioning; and following calibration users have complete freedom to move their heads. The resulting output is a volumetric cursor which can be used for advanced interfaces to allow direct 3D interaction with the world rather than via a computer VDU. We present the system as an alternative and complement to direct brain read out by BMIs. A simple performance metric to compare different BMIs is the information throughput—the rate at which the BMI communication interface can decode information from the brain. We calculate the theoretical information throughput achievable with our system and then compare it to information throughputs presented in an extensive review of BMIs [21]. The throughput is calculated as the product of bits communicated per unit command, and the number of unit commands that can be made per second. In the context of our gaze interface, each fixation can be considered as a unit command. With a sensory estimation error of 1.1 cm in width, 1.2 cm in height and 5.1 cm in depth (mean absolute error), over a workspace of 47 cm × 27 cm × 108 cm (width × height × depth),

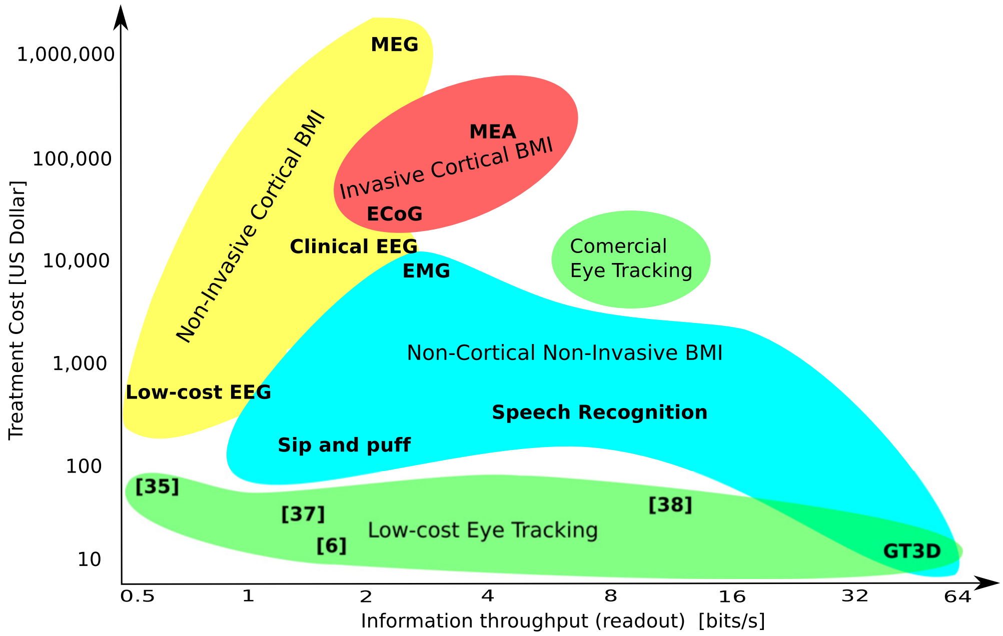

Figure 6. Comparison of different BMI and eye tracking technologies in terms of their treatment and hardware costs (in USD) and readout performance (measured as bits/s). Note, we used a log10 scale for the treatment cost and binary logarithm scale for the bit rate. The bit-rate data invasive and non-invasive BMIs were taken from [21], except stated otherwise. Treatment costs were taken from published data that were available (cited below), or from quotes we directly obtained from manufacturers and healthcare providers. GT3D—our system (component cost). EMG—electromyography (cost based on g.Hiamp EMG kit; Guger Technologies, Schiedlberg, Austria). 'Sip and puff'—switches actuated by user inhaling or exhaling (system cost from www.liberator.co.uk). Speech recognition—speech actuated commands (cost based on commercial speech recognition system Dragon's 'Naturally Speaking Software'). MEG—magnetoencephalography [32]. EEG—electroencephalography; clinical EEG (cost based on g.BCI EEG kit, Guger Technologies Gmbh, Schiedlberg, Austria), low-cost EEG (Emotive EEG headset kit, Emotiv, San Francisco, CA), bit rate from [33]. ECoG—electrocorticography; MEA—multielectrode array, cost of clinical research systems is based on Utah electrode arrays (Blackrock Systems, Salt Lake City, UT) and peripheral equipment plus the preoperative assessment, surgery, postoperative management cost estimated from deep brain stimulation costs [34]. Commercial eye tracking costs for 2D gaze tracking (Eyelink II, SR Research, Kanata, Ontario) with bit rate reported in [21]. Low-cost eye tracking—citations for individual prototype systems and their reported bit rates ([6]; [35] bit rate based on 40 characters per second text writing performance times 1 bit entropy per character of English language [36] yielding 0.67 bits s^{-1} [37]). The system recognizes ten different gaze gestures with an average of 2.5 s per gesture, yielding 1.3 bits s^{-1} [38]; the system recognizes 16 different gaze states at 3 states per second (average number of fixations per second) yielding 12 bits s^{-1}. Note: all bit rates reflect published values, higher read-out rates may be possible with other decoding strategies. Commercial device cost based on in-production items including ethical licenses. Low-cost eye tracking costs are prototype hardware costs.

there are 2.04 × 10^4 distinguishable states giving 14.3 bits of information per fixation. On average, we fixate with a rate of three fixations per second [22]; giving a bit rate of 43 bits s^{-1}. Our theoretical upper limit is significantly higher than other BMI mechanisms and the signal is obtained non-invasively, for a significantly lower cost. The information throughput reflects the accuracy of our gaze-controlled real-time continuous volumetric cursor, which yields a fast control signal with very low latency. Both the speed of information transmission, but also the natural role of gaze in attention and actions make our system highly suitable for controlling disability aids such as electric wheelchairs or end-points of prosthetic arms. We envisage the user tracing out their desired path using their eyes or looking at an object they wish to grasp and then guiding the object's manipulation. Figure 6 demonstrates the relationship between estimated treatment cost and bit-rates for different BMI mechanisms, including our system, labelled as GT3D in the plot. The treatment costs are estimated based on device cost as well as operational setup and maintenance costs such as surgery and rehabilitation costs (see also figure 6). The information rates given in [21] may underestimate throughput capacities of the different BMI methods, but at least offer a basic consistent benchmark across different readout technologies. Our system has an estimated information transfer capacity of 43 bits s^{-1}, which is ten times higher than other invasive BMI approaches (see figure 6), with closed-loop response latencies (measured from eye movement to computer response) below 10 ms.

BMI information rates from direct recording of neuronal activity are ultimately constrained by noise in the recording systems and the nervous system itself [23]. In particular, physical noise sources inside central neurons [24, 25] and peripheral axons [25, 26] will limit decoding performance from limited numbers of independent neuronal sources. Thus to compensate for noise, signal decoders have to observe signals for longer periods of time, thereby increasing response latencies for direct BMIs at the moment. While these issues will be ameliorated by the steady progress of sensor quality and density [27], eye movements already offer a highly accurate, low-latency (and low cost) read out. This is because the brain has already evolved to minimize the role of noise and delays in eye movements, which form an aggregated output of the nervous system. The leap in readout performance (in terms of readout performance and latency) enables closed-loop real-time control of rehabilitative and domotic devices beyond what is achievable by current BMIs: for example it was estimated that powered wheelchair control requires, on average, 15.3 bits s^{-1} and full-finger hand prosthetics require 54.2 bits s^{-1} [21]. Our system demonstrated a clear improvement on low-level measures of BMI performance, but such technical measures mask the complexities of learning to use and operating BMIs in the clinic and daily-life. Therefore, we also introduce a real-world, closed-loop control benchmark—playing an arcade video game—as a high-level, behaviour-based measure for BMI performance. We reported the performance for both normal (mouse-based) use and using our GT3D system in our subject study to establish the benchmark and make its software available to the community. On average, naïve users of our gaze-based system achieved a game score within 12.5% of their own score when playing the game directly with a computer mouse, demonstrating that subjects achieved near-normal closed-loop real-time control. This is also reflected in the mean number of successfully returned shots per game using our system (average of 43 against the computer-mouse score of 53) despite the novel control modality and very short training time (10 min from first use). We, thus, demonstrated how ultra-low cost, non-invasive eye-tracking approach can form the basis of a real-time control interface for rehabilitative devices, making it a low-cost complement or alternative to existing BMI technologies.

Eye movements are vital for motor planning; we look where we are going, reaching and steering [28], and therefore 3D gaze information is highly correlated with user intentions in the context of navigation and manipulation of our surroundings [29, 30]. Our approach, unlike other BMI technologies, enables us to use gaze information to infer user intention in the context of its natural occurrence, e.g. steering a wheel chair 'by eye' gaze, as we are already looking where we are going. This approach drastically reduces training time and boost patient's adherence. Moreover, the structured statistics of human eye movements in real-world tasks enables us to build Bayesian decoders to further boost decoding accuracy, reliability and speed even in complex environments [31]. Our 3D gaze tracking approach lends itself ideally to complement, or when treatment costs are at a premium even replace, conventional BMI approaches.

Acknowledgments

The authors acknowledge financial support from the EPSRC. Early results have been presented as a TED talk in May 2010 (London, UK), as a poster presentation at CNS 2011 (Stockholm, Sweden) [39] and as a TEDx Imperial talk in February 2012 (http://www.youtube.com/watch?v=6ZteM65b76Y). They acknowledge the early technical contributions of Ian Beer, Aaron Berk, Oliver Rogers and Timothy Treglown for a Linux and 2D version of the system.

References

[1] Hochberg L, Nurmikko A and Donoghue J E 2012 Brain machine interface *Annu. Rev. Biomed. Eng.* **14** at press

[2] Kaminski H J, Richmonds C R, Kusner L L and Mitsumoto H 2002 Differential susceptibility of the ocular motor system to disease *Ann. New York Acad. Sci.* **956** 42–54

[3] Kaminski H J, Al-Hakim M, Leigh R J, Bashar M K and Ruff R L 1992 Extraocular muscles are spared in advanced Duchenne dystrophy *Ann. Neurol.* **32** 586–88

[4] Jordansen I K, Boedeker S, Donegan M, Oosthuizen L, di Girolamo M and Hansen J P 2005 D7.2 Report on a market study and demographics of user population *Communication by Gaze Interaction (COGAIN)* IST-2003-511598 (available at http://www.cogain.org/results/reports/COGAIN-D7.2.pdf)

[5] Schneider N, Bex P, Barth E and Dorr M 2011 An open-source low-cost eye-tracking system for portable real-time and offline tracking *Proc. 1st Conf. on Novel Gaze-Controlled Applications (Karlskrona Sweden: ACM)* pp 1–4

[6] San Agustin J, Skovsgaard H, Hansen J P and Hansen D W 2009 Low-cost gaze interaction: ready to deliver the promises *Proc. 27th Int. Conf. Extended Abstracts on Human Factors in Computing Systems. CHI EA '09* (New York: ACM) pp 4453–58

[7] Li D, Babcock J and Parkhurst D J 2006 OpenEyes: a low-cost head-mounted eye-tracking solution *Proc. Symp. Eye Tracking Research and Applications. ETRA '06 (New York: ACM)* pp 95–100

[8] Jacob R J K 1990 What you look at is what you get: eye movement-based interaction techniques *Proc. SIGCHI Conf. on Human Factors in Computing Systems: Empowering People. CHI '90 (New York: ACM)* pp 11–8

[9] Duchowski A T, Medlin E, Cournia N and Gramopadhye A 2002 3D eye movement analysis for VR visual inspection training *Proc. Symp. Eye tracking research & applications. ETRA '02 (New York: ACM)* pp 103–10

[10] Ki J and Kwon Y-M 2008 3D gaze estimation and interaction *Conf. on 3DTV: The True Vision-Capture, Transmission and Display of 3D Video* pp 373–6

[11] Hennessey C and Lawrence P 2009 Noncontact binocular eye-gaze tracking for point-of-gaze estimation in three dimensions *IEEE Trans. Biomed. Eng.* **56** 790–99

[12] Sliney D H and Freasier B C 1973 Evaluation of optical radiation hazards *App. Opt.* **12** 1–24

[13] Sliney D *et al* 2005 Adjustment of guidelines for exposure of the eye to optical radiation from ocular instruments: statement from a task group of the International Commission on Non-Ionizing Radiation Protection (ICNIRP) *App. Opt.* **44** 2162–76

[14] Lee E C and Park K R 2008 A robust eye gaze tracking method based on a virtual eyeball model *Mach. Vis. Appl.* **20** 319–37

[15] Sheng-Wen S and Jin L 2004 A novel approach to 3D gaze tracking using stereo cameras *IEEE Trans. Syst. Man Cybern.* B **34** 234–45

[16] Sheng-Wen S, Yu-Te W and Jin L 2000 A calibration-free gaze tracking technique *Proc. 15th Int. Conf. on Pattern Recogn. (IEEE)* 4 pp 201–4

[17] Hennessey C A and Lawrence P D 2009 Improving the accuracy and reliability of remote system-calibration-free eye-gaze tracking *IEEE Trans. Biomed. Eng.* **56** 1891–900

[18] Faisal A A, Fislage M, Pomplun M, Rae R and Ritter H 1998 Observation of human eye movements to simulate visual exploration of complex scenes *Technical Report* University of Bielefeld http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.38.8855

[19] Morimoto C H and Mimica M R M 2005 Eye gaze tracking techniques for interactive applications *Comput. Vis. Image Underst.* **98** 4–24

[20] Brolly X L C and Mulligan J B 2004 Implicit calibration of a remote raze tracker *Proc. Int. Conf. Computer Vision and*

Pattern Recognition Workshop CVPRW '04 (Washington DC: IEEE Comp. Soc.) vol 8 p 134

[21] Tonet O, Marinelli M, Citi L, Rossini P M, Rossini L, Megali G and Dario P 2008 Defining brain-machine interface applications by matching interface performance with device requirements *J. Neurosci. Methods* **167** 91–104

[22] Land M F and Tatler B W 2009 *Looking and Acting: Vision and Eye Movements in Natural Behaviour* (New York: Oxford University Press)

[23] Faisal A A, Selen L P J and Wolpert D M 2008 Noise in the nervous system *Nature Rev. Neurosci.* **9** 292–303

[24] Faisal A A, Laughlin S B and White J A 2002 How reliable is the connectivity in cortical neural networks? *Proc. IEEE Int. J. Conf. on Neural Networks '02 (IJCNN)* pp 1661–6 (available at http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1007767&contentType=Conference+Publications&sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A21694%29)

[25] Faisal A 2010 Stochastic simulation of neurons, axons and action potentials *Stochastic Methods in Neuroscience* (Oxford: Oxford University Press) pp 297–343

[26] Faisal A A, White J A and Laughlin S B 2005 Ion-channel noise places limits on the miniaturization of the brain's wiring *Curr. Biol.* **15** 1143–9

[27] Stevenson I H and Kording K P 2011 How advances in neural recording affect data analysis *Nature Neurosci.* **14** 139–42

[28] Land M F and Lee D N 1994 Where we look when we steer *Nature* **6483** 742

[29] Hwang A D, Wang H-C and Pomplun M 2011 Semantic guidance of eye movements in real-world scenes *Vis. Res.* **51** 1192–205

[30] Land M F, Mennie N and Rusted J 1999 The roles of vision and eye movements in the control of activities of daily living *Perception* **28** 1311–28

[31] Wang H-C, Hwang A D and Pomplun M 2010 Object frequency and predictability effects on eye fixation durations in real-world scene viewing *J. Eye Mov. Res.* **3** 1–10

[32] Ray A and Bowyer S M 2010 Clinical applications of magnetoencephalography in epilepsy *Ann. Indian Acad. Neurol.* **13** 14–22

[33] Bobrov P, Frolov A, Cantor C, Fedulova I, Bakhnyan M and Zhavoronkov A 2011 Brain-computer interface based on generation of visual images *PLoS One* **6** e20674

[34] McIntosh E, Gray A and Aziz T 2003 Estimating the costs of surgical innovations: the case for subthalamic nucleus stimulation in the treatment of advanced Parkinson's disease *Mov. Disord.* **18** 993–9

[35] Lemahieu W and Wyns B 2011 Low cost eye tracking for human-machine interfacing *J. Eyetracking, Vis. Cogn. Emotion* **1** 1–12

[36] Shannon C 1951 *Prediction and Entropy of Printed English (Shannon: Collected Papers)* (Piscataway, NJ: IEEE)

[37] Rozado D, Rodriguez F B and Varona P 2012 Low cost remote gaze gesture recognition in real time *Appl. Soft Comput.* **12** 2072–84

[38] Topal C, Gerek N and Dogan A 2008 A head-mounted sensor-based eye tracking device: eye touch system *Proc. 2008 Symp. on Eye Tracking Research & Applications (ETRA '08) ACM (New York, NY, USA)* pp 87–90

[39] Abbott W W and Faisal A A 2011 Ultra-low cost eyetracking as an high-information throughput alternative to BMIs *BMC Neurosci.* **12** 103